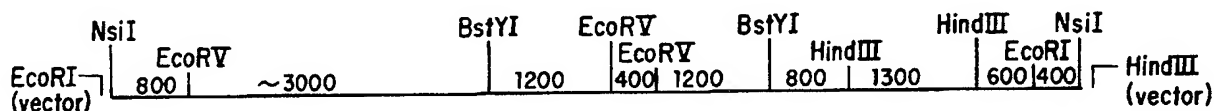


PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 1/21, 9/88, 15/31, 15/60, 15/61		A1	(11) International Publication Number: WO 94/25567
			(43) International Publication Date: 10 November 1994 (10.11.94)
(21) International Application Number: PCT/US94/04495 (22) International Filing Date: 22 April 1994 (22.04.94) (30) Priority Data: 08/052,206 23 April 1993 (23.04.93) US 08/053,615 23 April 1993 (23.04.93) US (71) Applicant: AMERICAN CYANAMID COMPANY [US/US]; One Cyanamid Plaza, Wayne, NJ 07470 (US). (72) Inventors: RYAN, Michael, Joseph; 34 Northwood Drive, West Milford, NJ 07480 (US). KHANDKE, Kiran, Manohar; 120 East Allison Avenue, Nanuet, NY 10954 (US). TILLEY, Bruce, Clifford; RR2, Box 27, Cresco, PA 18326 (US). LOTVIN, Jason, Arnold; 985 Roosevelt Avenue, Union, NJ 07083 (US). (74) Agent: GORDON, Alan, M.; American Cyanamid Company, One Cyanamid Plaza, Wayne, NJ 07470 (US).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>	

(54) Title: CLONING AND EXPRESSION OF THE CHONDROITINASE I AND II GENES FROM *P. VULGARIS*

(57) Abstract

This invention relates to the DNA sequence encoding the major protein component of chondroitinase ABC, which is referred to as "chondroitinase I", from *Proteus vulgaris* (*P. vulgaris*), which is contained in the Nsi fragment shown in the figure. This invention further relates to the DNA sequence encoding a second protein component of chondroitinase ABC, which is referred to as "chondroitinase II", from *P. vulgaris*, to the cloning and expression of the genes containing these DNA sequences, to the amino acid sequences of the recombinant chondroitinase I and II, and to methods for the isolation and purification of recombinant chondroitinase I or II. These methods provide significantly higher yields and purity than those obtained by adapting for the recombinant enzymes the method previously used for isolating and purifying native chondroitinase I enzyme from *P. vulgaris*.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

CLONING AND EXPRESSION OF THE
CHONDROITINASE I AND II GENES FROM P. VULGARIS

Field of the Invention

This invention relates to the DNA sequence encoding the major protein component of chondroitinase ABC, which is referred to as "chondroitinase I", from Proteus vulgaris (P. vulgaris). This invention
5 further relates to the DNA sequence encoding a second protein component of chondroitinase ABC, which is referred to as "chondroitinase II", from P. vulgaris. This invention also relates to the cloning and expression of the genes containing these DNA sequences
10 and to the amino acid sequences of the recombinant chondroitinase I and II enzymes encoded by these DNA sequences.

This invention additionally relates to methods for the isolation and purification of the
15 recombinantly expressed major protein component of chondroitinase ABC, which is referred to as "chondroitinase I", from Proteus vulgaris (P. vulgaris). This invention further relates to methods for the isolation and purification of the
20 recombinantly expressed second protein component of chondroitinase ABC, which is referred to as "chondroitinase II", from P. vulgaris. These methods provide significantly higher yields and purity than those obtained by adapting for the recombinant enzymes
25 the method previously used for isolating and purifying the native chondroitinase I enzyme from P. vulgaris.

- 2 -

Background of the Invention

Chondroitinases are enzymes of bacterial origin which have been described as having value in dissolving the cartilage of herniated discs without disturbing the stabilizing collagen components of those discs.

Examples of chondroitinase enzymes are chondroitinase ABC, which is produced by the bacterium P. vulgaris, and chondroitinase AC, which is produced by A. aurescens. The chondroitinases function by degrading polysaccharide side chains in protein-polysaccharide complexes, without degrading the protein core.

Yamagata et al. describes the purification of the enzyme chondroitinase ABC from extracts of P. vulgaris (Bibliography entry 1). The enzyme selectively degrades the glycosaminoglycans chondroitin-4-sulfate, dermatan sulfate and chondroitin-6-sulfate (also referred to respectively as chondroitin sulfates A, B and C) at pH 8 at higher rates than chondroitin or hyaluronic acid. However, the enzyme did not attack keratosulfate, heparin or heparitin sulfate.

Kikuchi et al. describes the purification of glycosaminoglycan degrading enzymes, such as chondroitinase ABC, by fractionating the enzymes by adsorbing a solution containing the enzymes onto an insoluble sulfated polysaccharide carrier and then desorbing the individual enzymes from the carrier (2).

Brown describes a method for treating intervertebral disc displacement in mammals, including humans, by injecting into the intervertebral disc space effective amounts of a solution containing chondroitinase ABC (3). The chondroitinase ABC was

- 3 -

isolated and purified from extracts of P. vulgaris. This native enzyme material functioned to dissolve cartilage, such as herniated spinal discs. Specifically, the enzyme causes the selective
5 chemonucleolysis of the nucleus pulposus which contains proteoglycans and randomly dispersed collagen fibers.

Hageman describes an ophthalmic vitrectomy method for selectively and completely disinserting the
10 ocular vitreous body, epiretinal membranes or fibrocellular membranes from the neural retina, ciliary epithelium and posterior lens surface of the mammalian eye as an adjunct to vitrectomy, by administering to the eye an effective amount of an
15 enzyme which disrupts or degrades chondroitin sulfate proteoglycan localized specifically to sites of vitreoretinal adhesion and thereby permit complete disinsertion of said vitreous body and/or epiretinal membranes (4). The enzyme can be a protease-free
20 glycosaminoglycanase, such as chondroitinase ABC. Hageman utilized chondroitinase ABC obtained from Seikagaku Kogyo Co., Ltd., Tokyo, Japan.

In isolating and purifying the chondroitinase ABC enzyme from the Seikagaku Kogyo
25 material, it was noted that there was a correlation between effective preparations of the chondroitinase in vitrectomy procedures and the presence of a second protein having an apparent molecular weight (by SDS-PAGE) slightly greater than that of the major protein
30 component of chondroitinase ABC. The second protein is now designated "chondroitinase II", while the major protein component of chondroitinase ABC is referred to as "chondroitinase I." The chondroitinase I and II proteins are basic proteins at neutral pH, with
35 similar isoelectric points of 8.30-8.45. Separate

- 4 -

purification of the chondroitinase I and II forms of the native enzyme revealed that it was the combination of the two proteins that was active in the surgical vitrectomy rather than either of the proteins individually.

5 Use of the chondroitinase I and II forms of the native enzyme to date has been limited by the small amounts of enzymes obtained from native sources. The production and purification of the native forms of the enzyme has been carried out using fermentations of P. vulgaris in which its substrate has been used as the inducer to initiate production of these forms of the enzyme. A combination of factors, including low levels of synthesis, the cost and availability of the inducer (chondroitin sulfate), and the opportunistically pathogenic nature of P. vulgaris, has resulted in the requirement for a more efficient method of production. In addition, the native forms of the enzyme produced by conventional techniques are subject to degradation by proteases present in the bacterial extract. Therefore, there is a need for a reliable supply of pure material free of contaminants in order for the medical applications of the two forms of this enzyme to be evaluated properly and exploited. There is also a need for methods to isolate and purify a reliable supply of the chondroitinase I and II enzymes free of contaminants.

Summary of the Invention

30 Accordingly, it is an object of this invention to produce chondroitinase I and chondroitinase II in quantities not readily achievable using present non-recombinant bacterial fermentation and extraction techniques.

35

- 5 -

It is a further object of this invention to produce chondroitinase I and chondroitinase II, each in a form substantially free of proteases which would otherwise degrade the enzyme and cause a loss of its activity.

These objects are achieved through the use of an alternative approach to the problems presented by large scale bacterial fermentation of these two forms of the enzyme. Separately for chondroitinase I and chondroitinase II, the gene that encodes the enzyme is cloned and the enzyme is expressed at high levels in a heterologous host. In a preferred embodiment, this invention is directed to the cloning of the P. vulgaris gene for chondroitinase I and the high level expression of that enzyme in E. coli, as well as the cloning of the P. vulgaris gene for chondroitinase II and the high level expression of that enzyme in E. coli.

This invention provides a purified isolated DNA fragment of P. vulgaris which comprises a sequence encoding for chondroitinase I. This invention further provides a purified isolated DNA fragment of P. vulgaris which hybridizes with a nucleic acid sequence encoding for amino acids as follows:

- (a) the chondroitinase I enzyme with its signal peptide (SEQ ID NO:2, amino acids 1-1021) or a biological equivalent thereof (encoded for example by: (1) nucleotides numbered 119-3181 of SEQ ID NO:1, and (2) nucleotides numbered 119-3181 of SEQ ID NO:3, where the three nucleotides immediately upstream of the initiation codon are changed (SEQ ID NO:3, nucleotides 116-118));

- 6 -

- (b) the mature chondroitinase I enzyme (SEQ ID NO:2, amino acids 25-1021) or a biological equivalent thereof (encoded for example by: (1) nucleotides
5 numbered 191-3181 of SEQ ID NO:1, and
(2) nucleotides numbered 191-3181 of
SEQ ID NO:3, where the three
nucleotides immediately upstream of the
initiation codon are changed (SEQ ID
10 NO:3, nucleotides 116-118)); and
- (c) the mature chondroitinase I enzyme
where the sequence encoding the signal
peptide has been replaced with a
sequence which adds a methionine
15 residue to the amino terminus of the
enzyme (SEQ ID NO:5, amino acids 24-
1021) or a biological equivalent
thereof (encoded for example by
nucleotides numbered 188-3181 of SEQ ID
20 NO:4).

The recombinant chondroitinase I is produced
by transforming a host cell with a plasmid containing
a purified isolated DNA fragment of P. vulgaris which
contains one of the above-described sequences, and
25 culturing the host cell under conditions which permit
expression of the enzyme by the host cell.

This invention also provides a purified
isolated DNA fragment of P. vulgaris which comprises a
sequence encoding for chondroitinase II. This
30 invention further provides a purified isolated DNA
fragment from P. vulgaris which hybridizes with a
nucleic acid sequence encoding for amino acids as
follows:

- (a) the chondroitinase II enzyme with its
35 signal peptide (SEQ ID NO:40, amino

- 7 -

acids 1-1013) or a biological equivalent thereof (encoded for example by nucleotides numbered 3238-6276 of SEQ ID NO:39); and

- 5 (b) the mature chondroitinase II enzyme (SEQ ID NO:40, amino acids 24-1013) or a biological equivalent thereof (encoded for example by nucleotides numbered 3307-6276 of SEQ ID NO:39).

10 The recombinant chondroitinase II is produced by transforming a host cell with a plasmid containing a purified isolated DNA fragment of P. vulgaris which contains one of the above-described sequences, and culturing the host cell under
15 conditions which permit expression of the enzyme by the host cell.

It is an additional object of this invention to provide methods for the isolation and purification of the recombinantly expressed chondroitinase I enzyme
20 of P. vulgaris.

It is a particular object of this invention to provide methods which result in significantly higher yields and purity of the recombinant chondroitinase I enzyme than those obtained by
25 adapting for the recombinant enzyme the method previously used for isolating and purifying the native chondroitinase I enzyme from P. vulgaris.

These objects are achieved through either of two methods described and claimed herein for the
30 chondroitinase I enzyme. The first method comprises the steps of:

- (a) lysing by homogenization the host cells which express the recombinant chondroitinase I enzyme to release the
35 enzyme into the supernatant;

- 8 -

- 5 (b) subjecting the supernatant to
diafiltration to remove salts and other
small molecules;
- (c) passing the supernatant through an
anion exchange resin-containing column;
- (d) loading the eluate from step (c) to a
cation exchange resin-containing column
so that the enzyme in the eluate binds
to the cation exchange column; and
- 10 (e) eluting the enzyme bound to the cation
exchange column with a solvent capable
of releasing the enzyme from the
column.

15 In the second method, prior to step (b) of
the first method just described, the following two
steps are performed:

- (1) treating the supernatant with an acidic
solution to precipitate out the enzyme;
and
- 20 (2) recovering the pellet and then
dissolving it in an alkali solution to
again place the enzyme in a basic
environment.

25 It is a further object of this invention to
provide methods for the isolation and purification of
the recombinantly expressed chondroitinase II enzyme
of P. vulgaris.

30 It is an additional object of this invention
to provide methods which result in significantly
higher yields and purity of the recombinant
chondroitinase II enzyme than those obtained by
adapting for the recombinant enzyme the method
previously used for isolating and purifying the native
chondroitinase I enzyme from P. vulgaris.

35 These objects are achieved through either of

- 9 -

two methods described and claimed herein for the chondroitinase II enzyme. The first method comprises the steps of:

- 5 (a) lysing by homogenization the host cells which express the recombinant chondroitinase I enzyme to release the enzyme into the supernatant;
- 10 (b) subjecting the supernatant to diafiltration to remove salts and other small molecules;
- (c) passing the supernatant through an anion exchange resin-containing column;
- (d) loading the eluate from step (c) to a cation exchange resin-containing column
15 so that the enzyme in the eluate binds to the cation exchange column;
- (e) obtaining by affinity elution the enzyme bound to the cation exchange column with a solution of chondroitin
20 sulfate, such that the enzyme is co-eluted with the chondroitin sulfate;
- (f) loading the eluate from step (e) to an anion exchange resin-containing column and eluting the enzyme with a solvent
25 such that the chondroitin sulfate binds to the column; and
- (g) concentrating the eluate from step (f) and crystallizing out the enzyme from the supernatant which contains an
30 approximately 37 kD contaminant.

In the second method, prior to step (b) of the first method just described, the following two steps are performed:

- 35 (1) treating the supernatant with an acidic solution to precipitate out the enzyme;

- 10 -

and

- (2) recovering the pellet and then dissolving it in an alkali solution to again place the enzyme in a basic environment.

5

Use of the methods of this invention results in significantly higher yields and purity of each recombinant enzyme than those obtained by adapting for each recombinant enzyme the method previously used for isolating and purifying the native chondroitinase I enzyme from P. vulgaris.

10

Brief Description of the Figures

15

Figure 1 depicts a preliminary restriction map for the subcloned approximately 10 kilobase Nsi fragment in pIBI24. The Nsi fragment contains the complete gene encoding chondroitinase I and a portion of the gene encoding chondroitinase II. The restriction sites are shown in their approximate positions. The restriction sites are useful in the constructions described below; other restriction sites present are not shown in this Figure; some are set forth in Example 13 below.

20

25

Figure 2 depicts the elution of the recombinant chondroitinase I enzyme from a cation exchange chromatography column using a sodium chloride gradient. The method used to purify the native enzyme is used here to attempt to purify the recombinant enzyme. The initial fractions at the left do not bind to the column. They contain the majority of the chondroitinase I enzyme activity. The fractions at right containing the enzyme are marked "eluted activity". The gradient is from 0.0 to 250 mM NaCl.

30

35

Figure 3 depicts the elution of the

- 11 -

recombinant chondroitinase I enzyme from a cation exchange column, after first passing the supernatant through an anion exchange column, in accordance with a method of this invention. The initial fractions at the left do not bind to the column, and contain only traces of chondroitinase I activity. The fractions at right containing the enzyme are marked "eluted activity". The gradient is from 0.0 to 250 mM NaCl.

Figure 4 depicts sodium dodecyl sulfate-polyacrylamide gel chromatography (SDS-PAGE) of the recombinant chondroitinase I enzyme before and after the purification methods of this invention are used. In the SDS-PAGE gel photograph, Lane 1 is the enzyme purified using the method of the first embodiment of the invention; Lane 2 is the enzyme purified using the method of the second embodiment of the invention; Lane 3 represents the supernatant from the host cell prior to purification -- many other proteins are present; Lane 4 represents the following molecular weight standards: 14.4 kD - lysozyme; 21.5 kD - trypsin inhibitor; 31 kD - carbonic anhydrase; 42.7 kD - ovalbumin; 66.2 kD - bovine serum albumin; 97.4 kD - phosphorylase B; 116 kD - beta-galactosidase; 200 kD - myosin. A single sharp band is seen in Lanes 1 and 2.

Figure 5 depicts SDS-PAGE chromatography of the recombinant chondroitinase II enzyme during various stages of purification using a method of this invention. In the SDS-PAGE gel photograph, Lane 1 is the crude supernatant after diafiltration; Lane 2 the eluate after passage of the supernatant through an anion exchange resin-containing column; Lane 3 is the enzyme after elution through a cation exchange resin-containing column; Lane 4 is the enzyme after elution through a second anion exchange resin-containing

- 12 -

column; Lane 5 represents the same molecular weight standards as described for Figure 4, plus 6.5 kD - aprotinin; Lane 6 is the same as Lane 4, except it is overloaded to show the approximately 37 kD
5 contaminant; Lane 7 is the 37 kD contaminant in the supernatant after crystallization of the chondroitinase II enzyme; Lane 8 is first wash of the crystals; Lane 9 is the second wash of the crystals; Lane 10 is the enzyme in the washed crystals after redissolving in water.

Detailed Description of the Invention

Preliminary experiments indicated that E. coli
15 coli could not use the hydrolysis products yielded by chondroitinase I as a sole carbon source, suggesting that this gene could not be cloned by selecting for its expression in E. coli. Another approach, followed in this application, is to use a physical method to
20 identify DNA fragments that encode the chondroitinase I enzyme. This is accomplished using an appropriately labeled probe for hybridization with individual clones that, together, make up a gene bank comprising the complete genome of P. vulgaris. The probe itself is
25 generated using Polymerase Chain Reaction (PCR) (5). In this procedure, the genomic DNA of P. vulgaris is denatured and oligonucleotides (designed to bracket part of the chondroitinase I gene) are annealed and DNA synthesis is carried out in vitro. This cycle of
30 denaturation, annealing and DNA synthesis using the oligonucleotides as primers is repeated many times (e.g., 30), with the yield of the desired product (the DNA fragment that lies between the two oligonucleotides) increasing exponentially with each cycle.

35 A putative nucleotide sequence of the

- 13 -

appropriate oligonucleotides is constructed from available amino acid sequence information derived from the protein purified from P. vulgaris bacteria. Once this is done, the DNA fragment produced by PCR is
5 cloned and its DNA sequence determined to verify that it is part of the chondroitinase I gene. It is then labeled and used as a probe to indicate which members of the gene bank actually contain the chondroitinase I gene. Subsequent restriction mapping and Southern
10 hybridization narrows the location to a piece of DNA of approximately four thousand base-pairs (bp). This is then sequenced using the Sanger dideoxy chain termination method (6) to reveal the exact position of the gene and guide the subsequent manipulations used
15 to place the gene into a high-level expression system in E. coli. A fermentation at a 10 liter scale carried out with this E. coli strain containing a recombinant plasmid expressing the P. vulgaris chondroitinase I gene yields a maximum chondroitinase
20 I titer of approximately 600 units/ml (which is the same as 1.2 mg/ml). This yield far exceeds that of the native P. vulgaris fermentation process which had not achieved a titer of more than 2 units/ml.

The process of cloning and expression of the
25 chondroitinase I gene is summarized by the following series of stages:

- 1) The isolation of P. vulgaris genomic DNA and the construction of a cosmid gene bank.
- 2) PCR experimentation designed to yield
30 an authentic piece of the chondroitinase I gene for use as a hybridization probe.
- 3) Colony hybridization studies to identify at least a portion of the chondroitinase I gene.
- 35 4) Restriction mapping, Southern hybridi-

- 14 -

zation, DNA sequencing, and chondroitinase I enzyme assays that, collectively, serve to place the location of the chondroitinase I gene more precisely within the cloned DNA.

5 5) DNA sequence analysis to reveal the exact coding region and location of the chondroitinase I gene.

10 6) Site-specific mutagenesis, related manipulations, and genetic engineering leading to the regulated, high-level expression of the P. vulgaris gene in E. coli.

15 These six stages are described in specific detail in Examples 1-7 below. The rationale for the stages is as follows. In the first stage, genomic DNA is obtained. DNA is separated from protein and other material contained in a P. vulgaris fermentation. Study of the genomic DNA is facilitated by the insertion of fragments of the DNA into cosmid vectors. The genomic DNA is digested with an appropriate
20 restriction endonuclease, such as Sau3A, and then ligated into a cosmid vector. The packaged recombinant cosmids containing the P. vulgaris DNA fragments are introduced into an appropriate bacterial host strain, such as an E. coli strain, and the
25 resulting culture is grown to allow gene expression. The gene banks are engineered to contain a marker, such as ampicillin or kanamycin resistance, to assist in the screening of the gene banks for the presence of the chondroitinase I gene.

30 Applicants have conducted some amino acid sequencing of the native chondroitinase I enzyme. Samples of the enzyme are generated by fermentation of P. vulgaris. Samples may also be obtained from Seikagaku Kogyo Co., Ltd., Tokyo, Japan. The amino
35 acid sequence information is used to design

- 15 -

oligonucleotides for use in screening for the chondroitinase I gene.

In the second stage, oligonucleotides are designed for use in PCR. A first set of
5 oligonucleotides is designed so as to encode a heptapeptide that has minimal degeneracy of its genetic code. Seven amino acids near the amino terminus of the chondroitinase I enzyme (SEQ ID NO:2, amino acids 19-25) are potentially encoded by 512
10 different nucleotide sequences (SEQ ID NO:6; see Example 2). The number of potential sequences is reduced to 32 by selecting specific nucleotides at the 5' end, because of the observation that mismatched nucleotides in PCR primers are of less consequence at
15 the 5' end than at the 3' end of the primer (7). The sequences of the pool of 32 primers are set out at SEQ ID NOS:7-14.

Applicants have discovered that the approximately 110 kD chondroitinase I enzyme is
20 cleaved proteolytically into an 18,000 MW ("18 kD") fragment and an approximately 90,000 MW ("90 kD") fragment. Furthermore, the 18 kD fragment is further fragmented by treatment with cyanogen bromide and trypsin. The various fragments are then used to
25 design additional sets of oligonucleotide primers for PCR.

Seven amino acids within the 18 kD fragment (SEQ ID NO:2, amino acids 114-120) are potentially encoded by 512 different nucleotide sequences (SEQ ID
30 NO:15; see Example 2). The complementary strand has the same number of potential sequences (SEQ ID NO:16; see Example 2). Using the criteria described above for the first set of oligonucleotides, the number of potential sequences is reduced to 128, whose sequences
35 are set out at SEQ ID NOS:17-24.

- 16 -

Six amino acids located near the amino-terminus of the "90 kD" fragment (SEQ ID NO:2, amino acids 165-170) are potentially encoded by a large number of different nucleotide sequences (SEQ ID NOS:25 and 26; see Example 2). The complementary strand has the same number of potential sequences (SEQ ID NOS:27 and 28; see Example 2). Using the criteria described above for the first set of oligonucleotides, the number of potential sequences is reduced to the sequences set out at SEQ ID NOS:29-36.

PCR amplifications are conducted using these 24 mixtures of oligonucleotides. The most effective amplifications are observed as discrete bands on electrophoretic gels. Products approximately 500 and 350 base pairs (bp) in size are obtained. The approximately 350 bp product is a subfragment of the approximately 500 bp product. The approximately 500 bp product is isolated and, following successive cloning procedures described in Example 2, is isolated as a 455 bp PCR product.

This 455 bp fragment is sequenced and translated into an amino acid sequence which is in virtual agreement with the sequence available from the native chondroitinase I enzyme. The sequences differ by one amino acid; subsequent experiments reveal that the nucleotide and amino acid sequences of the 455 bp fragment are correct, while the native amino acid sequence identification is in error.

In the third stage, the PCR amplification fragment is used as a probe to identify the cosmid gene banks prepared in the first stage which contain the chondroitinase I gene. The PCR fragment is denatured and labelled with, for example, digoxigenin-labelled dUTP (Boehringer-Mannheim, Indianapolis, IN). The cosmid gene banks are then used to infect a

- 17 -

bacterial strain. The resulting colonies are lysed and their DNA subjected to colony hybridization with the labelled probe, followed by exposure to an alkaline phosphatase-conjugated antibody to the digoxigenin-labelled material. Positive clones are visualized and then picked to be grown in selective media.

In the fourth stage, Southern hybridization (8) and restriction mapping are used to localize the position of the chondroitinase I gene within individual clones. The PCR-generated fragment described above is used as a Southern hybridization probe against *P. vulgaris* genomic DNA that is first digested by restriction enzymes and fractionated. In a second PCR amplification, several of the oligonucleotides described above are used as primers. The results indicate that the portion of the chondroitinase I gene that hybridizes to the probe is carried on several large DNA fragments.

These large DNA fragments are digested to yield individual fragments which are isolated, tested for the presence of chondroitinase I sequences by Southern hybridization, and then subcloned into appropriate vectors. Example 3 details the cloning strategy used. Restriction maps are generated to assist in the identification of the portions of the fragments carrying the desired sequences. In addition, *in vitro* chondroitinase I assays in which the activity of the enzyme based on measuring the release of unsaturated disaccharide from chondroitin sulfate C at 232 nm are conducted on several samples to assist in the placement and orientation of the chondroitinase I gene. The results of these procedures suggest that a 4.2 kb *EcoRV*-*EcoRI* fragment of a larger 10 kb *NsiI* fragment could contain the

- 18 -

entire chondroitinase I gene.

In the fifth stage, the above-mentioned 4.2 kb fragment is subjected to DNA sequence analysis. The resulting DNA sequence is 3980 nucleotides in length (SEQ ID NO:1). Translation of the DNA sequence into the putative amino acid sequence reveals a continuous open reading frame (SEQ ID NO:1, nucleotides 119-3181) encoding 1021 amino acids (SEQ ID NO:2).

In turn, analysis of the amino acid sequence reveals a 24 residue signal sequence (SEQ ID NO:2, amino acids 1-24), followed by a 997 residue mature (processed) chondroitinase I enzyme (SEQ ID NO:2, amino acids 25-1021).

Signal sequences are required for a complex series of post-translational processing steps which result in secretion of a protein from a host cell. The signal sequence constitutes the amino-terminal end of the protein to be secreted. In most cases, the signal sequence is cleaved off by a specific protease, called a signal peptidase.

The "18 kD" and "90 kD" fragments are found to be adjacent to each other, with the "18 kD" fragment constituting the first 157 amino acids of the mature protein (SEQ ID NO:2, amino acids 25-181), and the "90 kD" fragment constituting the remaining 840 amino acids of the mature protein (SEQ ID NO:2, amino acids 182-1021).

The chondroitinase I enzyme of this invention is expressed using established recombinant DNA methods. Suitable host organisms include bacteria, viruses, yeast, insect or mammalian cell lines, as well as other conventional organisms. The host cell is transformed with a plasmid containing a purified isolated DNA fragment encoding for

- 19 -

chondroitinase I enzyme. The host cell is then cultured under conditions which permit expression of the enzyme by the host cell.

5 In the sixth stage, the gene is subjected to site-directed mutagenesis to introduce unique restriction sites. These permit the gene to be moved, in the correct reading frame, into an expression system which results in expression of chondroitinase I enzyme at high levels. Such an appropriate host cell
10 is the bacterium E. coli.

As detailed in Example 6 below, two different constructs are prepared. In the first, the three nucleotides immediately upstream of the initiation codon are changed (SEQ ID NO:3, nucleotides
15 116-118) through the use of a mutagenic oligonucleotide (SEQ ID NO:37). The coding region and amino acid sequence encoded by the resulting construct are not changed, and the signal sequence is preserved (SEQ ID NO:3, nucleotides 119-3181; SEQ ID NO:2).

20 In a preferred embodiment of this invention, the second construct is used. In the second construct, the site-directed mutagenesis is carried out at the junction of the signal sequence and the start of the mature protein. A mutagenic
25 oligonucleotide (SEQ ID NO:38) is used which differs at six nucleotides from those of the native sequence (SEQ ID NO:1, nucleotides 185-190). The sequence differences result in (a) the deletion of the signal sequence, and (b) the addition of a methionine residue
30 at the amino-terminus, resulting in a 998 amino acid protein (SEQ ID NO:4, nucleotides 188-3181; SEQ ID NO:5).

In the absence of a signal sequence, the enzyme is not secreted. Fortunately, it is not
35 retained within the cell in the form of insoluble

- 20 -

inclusion bodies. Instead, at least some of the enzyme is produced intracellularly as a soluble active enzyme. The enzyme is extracted by homogenization, which serves to lyse the cells and thereby release the enzyme into the supernatant. Even with the signal sequence present, much of the enzyme is not secreted, because it is thought that this expression system provides such high yields of enzyme that it exceeds the capacity of the host cell to secrete that much enzyme.

As described in Example 7 below, the gene lacking the signal sequence is inserted into an appropriate expression vector. One such vector is pET-9A (9; Novagen, Madison, WI), which is derived from elements of the E. coli bacteriophage T7. The resulting recombinant plasmid is designated pTM49-6. The plasmid is then used to transform an appropriate expression host cell, such as the E. coli B strain BL21/(DE3)/pLysS (10; Novagen).

Samples of this E. coli B strain BL21(DE3)/pLysS carrying the recombinant plasmid pTM49-6 were deposited by Applicants on February 4, 1993, with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., and have been assigned ATCC accession number 69234.

Expression of the chondroitinase I enzyme using the deposited host cell yields approximately 300 times the amount of the enzyme as was possible using a same size fermentation vessel with native (non-recombinant) P. vulgaris.

After expression of the chondroitinase I enzyme, the supernatant from the host cells is treated to isolate and purify the enzyme. Initial attempts to isolate and purify the recombinant chondroitinase I enzyme do not result in high yields of purified

- 21 -

protein. The previous method for isolating and purifying native chondroitinase I from fermentation cultures of P. vulgaris is found to be inappropriate for the recombinant material.

5 The native enzyme is produced by fermentation of a culture of P. vulgaris. The bacterial cells are first recovered from the medium and resuspended in buffer. The cell suspension is then homogenized to lyse the bacterial cells. Then a
10 charged particulate such as Bioacryl (Toso Haas, Philadelphia, PA), is added to remove DNA, aggregates and debris from the homogenization step. Next, the solution is brought to 40% saturation of ammonium sulfate to precipitate out undesired proteins. The
15 chondroitinase I remains in solution.

The solution is then filtered and the retentate is washed to recover most of the enzyme. The filtrate is concentrated and subjected to diafiltration with a phosphate to remove the salt.

20 The filtrate containing the chondroitinase I is subjected to cation exchange chromatography using a cellulose sulfate column. At pH 7.2, 20 mM sodium phosphate, more than 98% of the chondroitinase I binds to the column. The native chondroitinase I is then
25 eluted from the column using a sodium chloride gradient.

The eluted enzyme is then subjected to additional chromatography steps, such as anion exchange and hydrophobic interaction column
30 chromatography. As a result of all of these procedures, chondroitinase I is obtained at a purity of 90-97%. The level of purity is measured by first performing SDS-PAGE. The proteins are stained using Coomassie blue, destained, and the lane on the gel is
35 scanned using a laser beam of wavelength 600 nm. The

- 22 -

purity is expressed as the percentage of the total absorbance accounted for by that band.

However, the yield of the native protein is only 25-35%. The yield is measured as the remaining activity in the final purified product, expressed as a percentage of the activity at the start (which is taken as 100%). In turn, the activity of the enzyme is based on measuring the release of unsaturated disaccharide from chondroitin sulfate C at 232 nm.

This purification method also results in the extensive cleavage of the approximately 110,000 dalton (110 kD) chondroitinase I protein into a 90 kD and an 18 kD fragment. Nonetheless, the two fragments remain non-covalently bound and exhibit chondroitinase I activity.

When this procedure is repeated with homogenate from lysed host cells carrying a recombinant plasmid encoding chondroitinase I, significantly poorer results are obtained. Less than 10% of the chondroitinase I binds to the cation exchange column at standard stringent conditions of pH 7.2, 20 mM sodium phosphate.

Under less stringent binding conditions of pH 6.8 and 5 mM phosphate, an improvement of binding with one batch of material to 60-90% is observed. However, elution of the recombinant protein with the NaCl gradient gives a broad activity peak, rather than a sharp peak (see Figure 2). This indicates the product is heterogeneous. Furthermore, in subsequent fermentation batches, the recombinant enzyme binds poorly (1-40%), even using the less stringent binding conditions. Most of these batches are not processed to the end, as there is poor binding. Therefore, their overall recovery is not quantified.

Based on these results, it is concluded that

- 23 -

the recombinant chondroitinase I enzyme has a reduced basicity compared to the native enzyme, and that the basicity also varies between batches, as well as within the same batch.

5 It is evident that the method used to isolate and purify the native enzyme is not appropriate for the recombinant enzyme. The method produces low yields of protein at high cost. Furthermore, for large batches, large amounts of
10 solvent waste are produced containing large amounts of a nitrogen-containing compound (ammonium sulfate). This is undesirable from an environmental point of view.

15 A hypothesis is then developed to explain these poor results and to provide a basis for developing improved isolation and purification methods. It is known that the native chondroitinase I enzyme is basic at neutral pH. It is therefore assumed that the surface of the enzyme has a net
20 excess of positive charges.

Without being bound by this hypothesis, it is believed that, in recombinant expression of the enzyme, the host cell contains or produces small, negatively charged molecules. These negatively
25 charged molecules bind to the enzyme, thereby reducing the number of positive charges on the enzyme. If these negatively charged molecules bind with high enough affinity to copurify with the enzyme, they can cause an alteration of the behavior of the enzyme on
30 the ion exchange column.

Support for this hypothesis is provided by the data described below. In general, cation exchange resins bind to proteins better at lower pH's than
35 higher pH's. Thus, a protein which is not very basic, and hence does not bind at a high pH, can be made to

- 24 -

bind to the cation exchanger by carrying out the operation at a lower pH. At pH 7.2, the native enzyme binds completely to a cation exchange resin. However, the recombinant-derived enzyme, due to the lowered basicity as a result of binding of the negatively charged molecules, does not bind very well (less than 10%). This enzyme can be made to bind up to 70% by using a pH of 6.8 and a lower phosphate concentration (5 mM rather than 20 mM), but heterogeneity and low yield remain great problems. Indeed, only one fermentation results in a 70% binding level; typically, it is much less (less than 10%) even at pH 6.8. This level of binding varies dramatically between different fermentation batches.

This hypothesis and a possible solution to the problem are then tested. If negatively charged molecules are attaching non-covalently to chondroitinase I, thus decreasing its basicity, it should be possible to remove these undesired molecules by using a strong, high capacity anion exchange resin. Removal of the negatively charged molecules should then restore the basicity of the enzyme. The enzyme could then be bound to a cation exchange resin and eluted therefrom in pure form at higher yields.

Experiments demonstrate that this approach indeed provides a solution to the problem encountered with the isolation and purification of the recombinantly expressed chondroitinase I enzyme.

As is discussed below, chondroitinase I is recombinantly expressed in two forms. The enzyme is expressed with a signal peptide, which is then cleaved to produce the mature enzyme. The enzyme is also expressed without a signal peptide, to produce directly the mature enzyme. The two embodiments of this invention which will now be discussed are

- 25 -

suitable for use in purifying either of these forms of the enzyme.

In the first embodiment of this aspect of the invention, the host cells which express the recombinant chondroitinase I enzyme are lysed by homogenization to release the enzyme into the supernatant. The supernatant is then subjected to diafiltration to remove salts and other small molecules. However, this step only removes the free, but not the bound form of the negatively charged molecules. The bound form of these charged species is next removed by passing the supernatant through a strong, high capacity anion exchange resin-containing column. An example of such a resin is the Macro-Prep™ High Q resin (Bio-Rad, Melville, N.Y.). Other strong, high capacity anion exchange columns are also suitable. Weak anion exchangers containing a diethylaminoethyl (DEAE) ligand also are suitable, although they are not as effective. Similarly, low capacity resins are also suitable, although they too are not as effective. The negatively charged molecules bind to the column, while the enzyme passes through the column. It is also found that some unrelated, undesirable proteins also bind to the column.

Next, the eluate from the anion exchange column is directly loaded to a cation exchange resin-containing column. Examples of such resins are the S-Sepharose™ (Pharmacia, Piscataway, N.J.) and the Macro-Prep™ High S (Bio-Rad). Each of these two resin-containing columns has SO_3^- ligands bound thereto in order to facilitate the exchange of cations. Other cation exchange columns are also suitable. The enzyme binds to the column and is then eluted with a solvent capable of releasing the enzyme from the column.

- 26 -

Any salt which increases the conductivity of the solution is suitable for elution. Examples of such salts include sodium salts, as well as potassium salts and ammonium salts. An aqueous sodium chloride solution of appropriate concentration is suitable. A
5 gradient, such as 0 to 250 mM sodium chloride is acceptable, as is a step elution using 200 mM sodium chloride.

10 A sharp peak is seen in the sodium chloride gradient elution (Figure 3). The improvement in enzyme yield over the prior method is striking. The recombinant chondroitinase I enzyme is recovered at a purity of 99% at a yield of 80-90%.

15 The purity of the protein is measured by scanning the bands in SDS-PAGE gels. A 4-20% gradient of acrylamide is used in the development of the gels. The band(s) in each lane of the gel is scanned using the procedure described above.

20 These improvements are related directly to the increase in binding of the enzyme to the cation exchange column which results from first using the anion exchange column. In comparative experiments, when only the cation exchange column is used, only 1% of the enzyme binds to the column. However, when the
25 anion exchange column is used first, over 95% of the enzyme binds to the column.

The high purity and yield obtained with the first embodiment of this invention make it more feasible to manufacture the recombinant chondroitinase
30 I enzyme on a large scale.

In a second embodiment of this aspect of the invention, two additional steps are inserted in the method before the diafiltration step of the first embodiment. The supernatant is treated with an acidic
35 solution to precipitate out the desired enzyme. The

- 27 -

pellet is recovered and then dissolved in an alkali solution to again place the enzyme in a basic environment. The solution is then subjected to the diafiltration and subsequent steps of the first embodiment of this invention.

In comparative experiments with the second embodiment of this invention, when only the cation exchange column is used, only 5% of the enzyme binds to the column. However, when the anion exchange column is used first, essentially 100% of the enzyme binds to the column. The second embodiment provides comparable enzyme purity and yield to the first embodiment of the invention.

Acid precipitation removes proteins that remain soluble; however, these proteins are removed anyway by the cation and anion exchange steps that follow (although smaller columns may be used). An advantage of the acid precipitation step is that the sample volume is decreased to about 20% of the original volume after dissolution, and hence can be handled more easily on a large scale. However, the additional acid precipitation and alkali dissolution steps of the second embodiment mean that the second embodiment is more time consuming than the first embodiment. On a manufacturing scale, the marginal improvements in purity and yield provided by the second embodiment may be outweighed by the simpler procedure of the first embodiment, which still provides highly pure chondroitinase I enzyme at high yields. An additional benefit of the two embodiments of the invention is that cleavage of the enzyme into 90 kD and 18 kD fragments is avoided.

The high purity of the enzyme produced by the two embodiments of this invention is depicted in Figure 4. A single sharp band is seen in the SDS-PAGE

- 28 -

gel photograph: Lane 1 is the enzyme using the method of the first embodiment; Lane 2 is the enzyme using the method of the second embodiment (Lane 3 represents the supernatant from the host cell prior to purification -- many other proteins are present; Lane 4 represents molecular weight standards).

The material deposited with the ATCC can also be used in conjunction with the sequences disclosed herein to regenerate the native chondroitinase I gene sequence (SEQ ID NO:1) or the modified chondroitinase I gene sequence which includes the signal sequence (SEQ ID NO:3) using conventional genetic engineering technology.

Production of native chondroitinase I enzyme in P. vulgaris after induction with chondroitin sulfate does not provide a high yield of enzyme; the enzyme represents approximately 0.1% of total protein present. When the recombinant construct with the signal sequence deleted is used in E. coli, approximately 15% of the total protein is the chondroitinase I enzyme.

In addition to the three DNA sequences just described for the chondroitinase I gene (SEQ ID NOS:1, 3 and 4), the present invention further comprises DNA sequences which, by virtue of the redundancy of the genetic code, are biologically equivalent to the sequences which encode for the enzyme, that is, these other DNA sequences are characterized by nucleotide sequences which differ from those set forth herein, but which encode an enzyme having the same amino acid sequences as those encoded by the DNA sequences set forth herein.

In particular, the invention contemplates those DNA sequences which are sufficiently duplicative of the sequences of SEQ ID NOS:1, 3 or 4 so as to

- 29 -

permit hybridization therewith under standard high stringency Southern hybridization conditions, such as those described in Sambrook et al. (11), as well as the biologically active enzymes produced thereby.

5 This invention also comprises DNA sequences which encode amino acid sequences which differ from those of the chondroitinase I enzyme, but which are the biological equivalent to those described for the enzyme (SEQ ID NOS:2 and 5). Such amino acid
10 sequences may be said to be biologically equivalent to those of the enzyme if their sequences differ only by minor deletions from or conservative substitutions to the enzyme sequence, such that the tertiary configurations of the sequences are essentially
15 unchanged from those of the enzyme.

 For example, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such
20 as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, as well as
25 changes based on similarities of residues in their hydropathic index, can also be expected to produce a biologically equivalent product. Nucleotide changes which result in alteration of the N-terminal or C-terminal portions of the protein molecule would also
30 not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Therefore, where the terms "chondroitinase
35 I gene" or "chondroitinase I enzyme" are used in

- 30 -

either the specification or the claims, each will be understood to encompass all such modifications and variations which result in the production of a biologically equivalent protein.

5 The starting point for the cloning and expression of the chondroitinase II gene is partial amino acid sequencing of the mature native chondroitinase II protein obtained from P. vulgaris. The N-terminal sequence of the mature native
10 chondroitinase II protein is found to include the following 22 amino acids:

Leu-Pro-Thr-Leu-Ser-His-Glu-Ala-Phe-Gly-Asp-Ile-Tyr-
15 Leu-Phe-Glu-Gly-Glu-Leu-Pro-Asn-Thr (SEQ ID NO: 40, amino acids 1-22)

 The nucleotide sequence determined above for the region encoding the chondroitinase I gene includes an additional approximately 800 base pairs beyond the
20 translation termination codon (SEQ ID NOS:1 and 39, nucleotides 3185-3980). An inspection of this region reveals that the sequence between nucleotides 3307 and 3372 (SEQ ID NOS:1 and 39) encodes the identical 22
25 amino acids in the same order as the first 22 amino acids of native chondroitinase II.

 Furthermore, an ATG initiation codon (SEQ ID NOS:1 and 39, nucleotides 3238-3240) is found upstream of this region and in-frame, indicating that this gene is expressed with a 23 amino acid signal peptide
30 sequence for the export of chondroitinase II (SEQ ID NO:40, amino acids 1-23). Although a Shine-Dalgarno sequence (AGGA; SEQ ID NOS:1 and 39, nucleotides 3225-3228) is found upstream of the initiation codon, there is no apparent promoter sequence, suggesting that both
35 the 110 kD and 112 kD forms of the P. vulgaris

- 31 -

chondroitinase enzyme are expressed as part of a single messenger RNA.

5 The coding sequence that starts with this ATG was originally not found to be continuous in SEQ ID NO:1, since a termination codon (TAA) was thought to be present in-frame at base-pairs identified as 3607-3609. Re-examination of the sequencing data, however, revealed that a residue was overlooked and that a T should be inserted between nucleotides
10 originally identified as 3593 and 3594. This change restores the open reading frame which then extends through the end of SEQ ID NO: 39 (SEQ ID NOS:1 and 39 include the inserted T as nucleotide 3594). (Thus, the three bases TAA at base-pairs 3608-3610, properly
15 numbered, do not constitute a termination codon.)

With this information available, the cloning and expression of the *P. vulgaris* chondroitinase II gene is performed in three stages. In the first stage, because the N-terminal sequences are known, a
20 site-specific mutagenesis is carried out. This is necessary in order for this gene to be placed, eventually, directly into the desired T7-based expression vector pET9A that is used (as described above) for the chondroitinase I gene. The mutagenized
25 bases are upstream of the coding region (an AT sequence (SEQ ID NOS:1 and 39, base pairs 3235 and 3236) is replaced by a CA sequence).

The second stage, which can be carried out in parallel with the first, involves the
30 identification, isolation and DNA sequencing of an appropriate DNA fragment which will include the C-terminal coding region of the chondroitinase II gene. The available DNA sequence information is adequate to account for approximately 220 amino acids of an
35 estimated 1000 for the entire chondroitinase II

- 32 -

protein. The missing coding sequences, therefore, would extend for another 2400 base pairs beyond the end of SEQ ID NO: 1.

5 The third stage involves the assembly of an intact gene for chondroitinase II that has been modified to include the initiation codon as part of an NdeI site and to be followed by a BamHI site downstream of the coding region. This allows a directed insertion of this gene into the pET9A
10 expression vector (Novagen, Madison, WI) without further modification.

Sequencing of the entire assembled gene confirms the presence of the initiation codon at nucleotides 3238-3240, where this codon represents the
15 start of the region coding for the signal peptides at nucleotides 3238-3306, the region coding for the mature protein at nucleotides 3307-6276, and a termination codon at nucleotides 6277-6279 (SEQ ID NO:39). The translation of this sequence results in
20 1013 amino acids, of which the first 23 amino acids are the signal peptide and 990 amino acids constitute the mature chondroitinase II protein at residues numbered 24-1013 (SEQ ID NO:40). In this construction, the signal peptide is retained, such
25 that the expressed gene is processed and secreted to yield the mature native enzyme structure that has a leucine residue at the N-terminus.

As described in Example 13 below, the gene encoding the chondroitinase II protein is inserted
30 into pET9A and the resulting recombinant plasmid is designated LP²1359. The plasmid is then used to transform an appropriate expression host cell, such as the E. coli B strain BL21(DE3)/pLysS (which is also used for the expression of the chondroitinase I gene.

35 Samples of this E. coli B strain designated

- 33 -

TD112, which is BL21(DE3)/pLysS carrying the recombinant plasmid LP²1359, were deposited by Applicants on April 6, 1994, with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., and have been assigned ATCC accession number 69598.

Expression of the chondroitinase II enzyme using the deposited host cell yields approximately 25 times the amount of the enzyme as was possible using a same size fermentation vessel with native (non-recombinant) P. vulgaris.

After expression of the enzyme, the supernatant from the host cells is treated to isolate and purify the enzyme. Because of the virtually identical isoelectric points and similar molecular weights for the two proteins, the first method described above for isolating and purifying the recombinant chondroitinase I protein is adapted for isolating and purifying the recombinant chondroitinase II protein, and then modified as will now be described.

The need for the modification of the method is based on the fact that the recombinant chondroitinase II protein is expressed at levels approximately several-fold lower than the recombinant chondroitinase I protein; therefore, a more powerful and selective solution is necessary in order to obtain a final chondroitinase II product of a purity equivalent to that obtained for the chondroitinase I protein.

The first several steps of the method for the chondroitinase II protein are the same as those used to isolate and purify the chondroitinase I protein. Initially, the host cells which express the recombinant chondroitinase II enzyme are lysed by

- 34 -

homogenization to release the enzyme into the supernatant. The supernatant is then subjected to diafiltration to remove salts and other small molecules. However, this step only removes the free, but not the bound form of the negatively charged molecules. The bound form of these charged species is next removed by passing the supernatant through a strong, high capacity anion exchange resin-containing column. An example of such a resin is the Macro-PrepTM High Q resin (Bio-Rad, Melville, N.Y.). Other strong, high capacity anion exchange columns are also suitable. Weak anion exchangers containing a diethylaminoethyl (DEAE) ligand also are suitable, although they are not as effective. Similarly, low capacity resins are also suitable, although they too are not as effective. The negatively charged molecules bind to the column, while the enzyme passes through the column. It is also found that some unrelated, undesirable proteins also bind to the column.

Next, the eluate from the anion exchange column is directly loaded to a cation exchange resin-containing column. Examples of such resins are the S-SepharoseTM (Pharmacia, Piscataway, N.J.) and the Macro-PrepTM High S (Bio-Rad). Each of these two resin-containing columns has SO_3^- ligands bound thereto in order to facilitate the exchange of cations. Other cation exchange columns are also suitable. The enzyme binds to the column, while a significant portion of contaminating proteins elute unbound.

At this point, the method diverges from that used for the chondroitinase I protein. Instead of eluting the protein with a non-specific salt solution capable of releasing the enzyme from the cation exchange column, a specific elution using a

- 35 -

solution containing chondroitin sulfate is used.

This procedure utilizes the affinity the positively charged chondroitinase II protein has for the negatively charged chondroitin sulfate. The affinity is larger than that accounted for by a simple positive and negative interaction alone. It is an enzyme-substrate interaction, which is similar to other specific biological interactions of high affinity, such as antigen-antibody, ligand-receptor, co-factor-protein and inhibitor/activator-protein. Hence, the chondroitin sulfate is able to elute the enzyme from the negatively charged resin. In contrast, the resin-enzyme interaction is a simple positive and negative interaction.

Although affinity elution chromatography is as easy to practice as ion-exchange chromatography, the elution is specific, unlike salt elution. Thus, it has the advantages of both affinity chromatography (specificity), as well as ion-exchange chromatography (low cost, ease of operation, reusability).

Another advantage is the low conductivity of the eluent (approximately 5% of that of the salt eluent), which allows for further ion-exchange chromatography without a diafiltration/dialysis step, which is required when a salt is used. Note, that this is not a consideration in the method for the chondroitinase I protein, because no further ion-exchange chromatography is needed in order to obtain the purified chondroitinase I protein.

There is another reason for not using the method for purifying recombinant chondroitinase I. Chondroitinase II obtained using the chondroitinase I salt elution purification method has poor stability; there is extensive degradation at 4°C within one week. In contrast, chondroitinase II obtained by affinity

- 36 -

elution is stable. The reason for this difference in stability is not known. It is to be noted that chondroitinase I obtained by salt elution is stable.

5 The cation exchange column is next washed with a phosphate buffer to elute unbound proteins, followed by washing with borate buffer to elute loosely bound contaminating proteins and to increase the pH of the resin to that required for the optimal elution of the chondroitinase II protein using the
10 substrate, chondroitin sulfate.

Next, a solution of chondroitin sulfate in water, adjusted to pH 9.0, is used to elute the chondroitinase II protein, as a sharp peak (recovery 65%) and at a high purity of approximately 95%. A 1%
15 concentration of chondroitin sulfate is used. A gradient of this solvent is also acceptable.

Because the chondroitin sulfate has an affinity for the chondroitinase II protein which is stronger than its affinity for the resin of the
20 column, the chondroitin sulfate co-elutes with the protein. This ensures that only protein which recognizes chondroitin sulfate is eluted, which is desirable, but also means that an additional process step is necessary to separate the chondroitin sulfate
25 from the chondroitinase II protein.

In this separation step, the eluate is adjusted to a neutral pH and is loaded as is onto an anion exchange resin-containing column, such as the Macro-Prep™ High Q resin. The column is washed with a
30 phosphate buffer. The chondroitin sulfate binds to the column, while the chondroitinase II protein flows through in the unbound pool with greater than 95% recovery. At this point, the protein is pure, except for the presence of a single minor contaminant of
35 approximately 37 kD. The contaminant may be a

- 37 -

breakdown product of the chondroitinase II protein.

This contaminant is effectively removed by a crystallization step. The eluate from the anion exchange column is concentrated and the solution is maintained at a reduced temperature, such as 4°C, for several days to crystallize out the pure chondroitinase II protein. The supernatant contains the 37 kD contaminant. Centrifugation causes the crystals to form a pellet, while the supernatant with the 37 kD contaminant is removed by pipetting. The crystals are then washed with water. The washed crystals are composed of the chondroitinase II protein at a purity of greater than 99%.

In a second embodiment of this aspect of the invention for the chondroitinase II protein, two additional steps are inserted in the method before the diafiltration step of the first embodiment. The supernatant is treated with an acidic solution to precipitate out the desired enzyme. The pellet is recovered and then dissolved in an alkali solution to again place the enzyme in a basic environment. The solution is then subjected to the diafiltration and subsequent steps of the first embodiment of this invention.

Acid precipitation removes proteins that remain soluble; however, these proteins are removed anyway by the cation and anion exchange steps that follow (although smaller columns may be used). An advantage of the acid precipitation step is that the sample volume is decreased compared to the original volume after dissolution, and hence can be handled more easily on a large scale. However, the additional acid precipitation and alkali dissolution steps of the second embodiment mean that the second embodiment is more time consuming than the first embodiment. On a

- 38 -

manufacturing scale, the marginal improvements in purity and yield provided by the second embodiment may be outweighed by the simpler procedure of the first embodiment, which still provides highly pure chondroitinase II enzyme at high yields.

5 Production of native chondroitinase II enzyme in P. vulgaris after induction with chondroitin sulfate does not provide a high yield of enzyme; the enzyme represents approximately 0.1% of total protein present. When the recombinant construct is used in E. coli, approximately 2.5% of the total protein is the chondroitinase II enzyme.

10 In addition to the DNA sequence just described for the chondroitinase II gene (SEQ ID NO:39), the present invention further comprises DNA sequences which, by virtue of the redundancy of the genetic code, are biologically equivalent to the sequences which encode for the enzyme, that is, these other DNA sequences are characterized by nucleotide sequences which differ from those set forth herein, but which encode an enzyme having the same amino acid sequences as those encoded by the DNA sequences set forth herein.

15 In particular, the invention contemplates those DNA sequences which are sufficiently duplicative of the sequence of SEQ ID NO:39 so as to permit hybridization therewith under standard high stringency Southern hybridization conditions, such as those described in Sambrook et al. (11), as well as the biologically active enzymes produced thereby.

20 This invention also comprises DNA sequences which encode amino acid sequences which differ from those of the chondroitinase II enzyme, but which are the biological equivalent to those described for the enzyme (SEQ ID NO:40). Such amino acid sequences may

25

30

35

- 39 -

be said to be biologically equivalent to those of the enzyme if their sequences differ only by minor deletions from or conservative substitutions to the enzyme sequence, such that the tertiary configurations of the sequences are essentially unchanged from those of the enzyme.

For example, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, as well as changes based on similarities of residues in their hydropathic index, can also be expected to produce a biologically equivalent product. Nucleotide changes which result in alteration of the N-terminal or C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Therefore, where the terms "chondroitinase II gene" or "chondroitinase II enzyme" are used in either the specification or the claims, each will be understood to encompass all such modifications and variations which result in the production of a biologically equivalent protein.

If desired, one of ordinary skill in the art can ligate together the two pieces of DNA from the two deposits, for example, at the HindIII site at nucleotide 3326, so as to express both the chondroitinase I and chondroitinase II proteins under

- 40 -

the control of the T7 promoter upstream of the coding sequence for chondroitinase I.

In order that this invention may be better understood, the following examples are set forth. The examples are for the purpose of illustration only and are not to be construed as limiting the scope of the invention.

Examples

Standard molecular biology techniques are utilized according to the protocols described in Sambrook et al. (11).

Example 1

Isolation Of *P. vulgaris* Genomic DNA And Construction Of A Cosmid Bank In *E. coli*

Two 35 ml aliquots (designated A and B) of a *P. vulgaris* large-scale (1000 liter) fermentation are obtained and centrifuged. Both pellets are resuspended with 7 ml of 0.05M glucose-0.025M Tris-HCl-0.01M EDTA (pH 8) containing 4 mg/ml of egg-white lysozyme. After 30 minutes of incubation at 37°C, 7 ml of 1% SDS-0.16M EDTA-0.02M NaCl (pH 8) are added to sample "A" and incubation is continued at 37°C for another hour.

After the initial lysozyme treatment, sample "B" is centrifuged and the cell pellet taken up with 7 ml of 0.05M glucose-0.025M Tris-HCl-0.01M EDTA (pH 8) containing 40 µg/ml of DNAase-free RNAase and then 7 ml of 1% SDS-0.16M EDTA-0.02M NaCl (pH 8) are added to this resuspended material. Finally, proteinase K (Boehringer Mannheim, Indianapolis, IN) is added to both samples to a final concentration of 100 µg/ml and

- 41 -

incubation is continued overnight at 37°C.

The next day, the samples are extracted once with an equal volume (14 ml) of equilibrated phenol followed by two further extractions in which the samples are extracted with 7 ml of phenol followed by the addition of 7 ml of chloroform, continued shaking and finally, centrifugation to separate the two phases. The DNA is precipitated by adding one-quarter volume of 5M ammonium acetate and 0.6 volumes of isopropanol followed by centrifugation. The pelleted DNA is rinsed once with 70% (v/v) ethanol, dried under vacuum and then resuspended with 1 ml of TE (0.01M Tris-HCl-0.001M EDTA, pH 7.4). The nucleic acid concentration for sample "A" is 1.2 mg/ml while that for sample "B" is 1.3 mg/ml, as determined by their ultraviolet absorption at 260 nm.

Fragmentation of the genomic DNA to yield pieces of a size suitable for insertion into cosmid vectors (approximately 25-35 kilobases (kb)) is accomplished by partial digestion with the restriction endonuclease Sau3A. Duplicate 0.2 ml reactions are set up (one with preparation "A" and the other with DNA from preparation "B"), each containing 100 µg of the P. vulgaris genomic DNA, 0.1M NaCl, 0.01M MgCl₂, 0.01M Tris-HCl (pH 7.5) and 80 units of the enzyme Sau3A.

Incubation is carried out at 37°C and 25 µl aliquots are removed at appropriate time points (5,6,7,8,9,10,11 and 20 minutes) and added to 25 µl of 0.2M EDTA (pH 8). The individual samples are heated to 70°C and then 10 µl are removed for a size-distribution analysis on an agarose gel. The sample obtained after five minutes of Sau3A digestion of preparation "A" and that obtained after 6 minutes with preparation "B" are chosen for further use.

- 42 -

In each case, an aliquot (4 μ l, which is approximately equal to 2 μ g) of the chosen partial digest is ligated to the appropriate "left" and "right" arms of the cosmid vector DNA using approximately 1 μ g and 2 μ g of each, respectively, in 10 μ l reactions containing 0.066M Tris-HCl (pH 7.4), 0.01M MgCl₂, 0.001M ATP, and 400 units (as defined by the manufacturer (New England Biolabs, Beverly, MA)) of T4 DNA ligase. Incubation is carried out at 11°C overnight. The "left" and "right" arms of the cosmids are DNA fragments which, when ligated to an appropriately sized piece of P. vulgaris DNA, comprise a recombinant molecule of approximately 35-50 kb. Both arms contain "cos" sites which are recognized by the packaging enzymes in the next step. In addition, these arms carry the origin of replication and ampicillin-resistance functions of pIBI24 (International Biochemical Inc., New Haven, CT).

Each of the above ligase reactions is added to one tube of a λ packaging extract (Packagene™, a trademark of Promega Corp., Madison, WI) and the reaction is allowed to proceed at room temperature for two hours, at which point 0.5 ml of PDB (0.1M NaCl-0.01M Tris-HCl (pH 7.9)-0.01M MgSO₄) is added followed by approximately 0.05 ml of chloroform. Each tube of packaged DNA is, therefore, a gene bank of the P. vulgaris genome.

Because this method of construction creates a pool of infectious particles (i.e., λ phage heads filled with the cosmid vector joined to approximately 25 to 35 kb of P. vulgaris DNA), the number of potential clones is quantitated by adsorbing an aliquot of the packaged material to an appropriate, sensitive E. coli host strain, and then after outgrowth, plating the mixture on selective media.

- 43 -

For example, an overnight culture of the E. coli strain ER1562 (New England Biolabs, Beverly, MA) grown in 20-10-5 medium is diluted 1:20 into fresh media (20-10-5 supplemented with 1% maltose) and grown for three hours at 37°C. The cells (1 ml) are then centrifuged, resuspended with PDB (0.2 ml) and 0.02 ml of the appropriate gene bank added. After adsorption for twenty minutes at 37°C, the samples are diluted to 2 ml with 20-10-5 medium and grown at 37°C for 30 minutes. The culture is then spread on 20-10-5 plates containing 100 µg/ml of ampicillin and colonies scored after overnight incubation at 37°C. The results indicate that there are approximately 68,000 and 95,000 infectious particles (potential cosmid clones) present in the two samples, designated PV1-GB and PV2-GB, corresponding to the "A" and "B" preparation of P. vulgaris genomic DNA, respectively.

In addition, four other P. vulgaris gene banks are prepared, as above, using two different cosmid vectors. These two cosmids differ from the above-mentioned vectors in that a kanamycin resistance determinant is used in one case rather than the ampicillin resistance, while in the other, the replication functions of pBR322 (New England Biolabs, Beverly, MA) are used instead of those of pIBI24. These four "libraries," designated L1974, L1975, L1976, and L1977, contain, respectively, approximately 18,000 (amp^r), 34,000 (amp^r), 13,000 (kan^r) and 15,000 (kan^r) members. Aliquots of each of these six gene banks are screened for the presence of the P. vulgaris chondroitinase I gene (see below).

- 44 -

Example 2

PCR Experimentation Designed To Yield An
Authentic Piece Of The Chondroitinase I Gene
For Use As A Hybridization Probe

5

The Polymerase Chain Reaction (PCR) (5)
allows the geometric amplification of a DNA sequence
that lies between oligonucleotide primers that can be
extended by a DNA polymerase in vitro. The enzyme
10 used in these experiments is the Taq DNA polymerase
(isolated originally from Thermus aquaticus), which is
preferred because of its thermotolerance which allows
it to survive the repeated DNA denaturation steps that
are carried out at 94°C.

15 In order for this method to be employed
successfully, the oligonucleotides used must have
sequences that are as close as possible to those of
the target sequence -- the P. vulgaris chondroitinase
I gene. An approximation of that sequence can be
20 derived from the limited available amino acid sequence
data. To minimize uncertainty in the sequence
presented by the degeneracy of the genetic code (a
given amino acid can be encoded by up to six codons),
the first approximation involves choosing an amino
25 acid sequence that has the least degeneracy. For
example, in the amino-terminal sequence of the P.
vulgaris chondroitinase I gene, there are the
following consecutive amino acids: His-Phe-Ala-Gln-
Asn-Asn-Pro (SEQ ID NO:2, amino acids 43-49).

30 This amino acid sequence could be encoded by
any one of 512 different nucleotide sequences, repre-
sented as 5'-CAY-TTY-GCN-CAR-AAY-AAY-CCN-3' (SEQ ID
NO:6), where R stands for purine (A or G), Y for
pyrimidine (C or T), and N indicates that any one of
35 the four nucleotides (A T, G, or C) at this position

- 45 -

will constitute a nucleotide sequence that could encode the indicated amino acid sequence. One possible approach would be to synthesize an oligonucleotide mixture containing a total of 512 different oligonucleotides, represented as:

5' -CA(TC) -TT(TC) -GC(GATC) -CA(GA) -AA(TC) -AA(TC) -CC-(GATC) -3' (SEQ ID NO:6).

Although use of such mixtures in PCR has been successful, another approach is to use a number of oligonucleotide mixtures, each of which is made up of a relatively smaller set of nucleotide sequences. In order to simplify this further, advantage is taken of the observation (7) that mismatched nucleotides in PCR primers are of less consequence at the 5'-end of the primer than they are at the 3'-end. Using these criteria, a set of eight oligonucleotides (each made up of four unique sequences) is designed, where the individual sets of oligonucleotides have the following sequences:

1. 5' -CAC-TTC-GC(GATC) -CAA-AAT-AAT-CC-3' (SEQ ID NO:7)
2. 5' -CAC-TTC-GC(GATC) -CAA-AAC-AAC-CC-3' (SEQ ID NO:8)
3. 5' -CAC-TTC-GC(GATC) -CAA-AAC-AAT-CC-3' (SEQ ID NO:9)
4. 5' -CAC-TTC-GC(GATC) -CAA-AAT-AAC-CC-3' (SEQ ID NO:10)
5. 5' -CAC-TTC-GC(GATC) -CAG-AAT-AAT-CC-3' (SEQ ID NO:11)
6. 5' -CAC-TTC-GC(GATC) -CAG-AAC-AAC-CC-3' (SEQ ID NO:12)
7. 5' -CAC-TTC-GC(GATC) -CAG-AAC-AAT-CC-3' (SEQ ID NO:13)
8. 5' -CAC-TTC-GC(GATC) -CAG-AAT-AAC-CC-3' (SEQ ID NO:14)

One of these pools is perfectly matched for the first eleven nucleotides (counting from the 3-end), and, furthermore, within this pool of four oligonucleotides, one is a perfect match for the first

- 46 -

fourteen nucleotides. This is important because it permits stringent annealing conditions to be used that discriminate against imperfect matches that give rise to PCR products that are unrelated to the chondroitinase I gene.

5 A further aid in the design of oligonucleotides to be used in these PCR experiments is derived from the observation that the *P. vulgaris* 110 kD chondroitinase enzyme appears to have a structure that leaves one particular region hypersensitive to
10 proteolytic cleavage. The result of this hydrolysis is that the normally approximately 110 kD protein is split into two predominant species of 18 kD and approximately 90 kD. The amino-terminal sequences of
15 the "110 kD" protein and the "18 kD" fragment are the same, while that for the "90 kD" has been found to be different.

The "18 kD" peptide is further fragmented by treatment with cyanogen bromide and trypsin and the
20 resulting oligopeptides sequenced, affording still more information with which to design oligonucleotides for PCR. This information from the "18 kD" and "90 kD" regions is also valuable because the locations of these amino acid sequences relative to each other and
25 the N-terminal sequences of the intact protein are well defined. In fact, the nucleotide distance between the regions encoding the N-termini of the "110 kD" and "90 kD" entities can be predicted to be approximately 400-500 bp.

30 Two further sets of oligonucleotide pools are then designed with one further consideration: The first eight oligonucleotides hybridize to one strand of the DNA and, during the *in vitro* DNA synthesis, they are extended toward the "90 kD" N-terminal coding
35 sequences. Consequently, the oligonucleotides

- 47 -

corresponding to amino acid sequences from within the "18 kD" peptide and at the N-terminus of the "90 kD" peptide must be designed so that they anneal to the complementary DNA strand of the P. vulgaris genome, so that they extend, in vitro, toward the region encoding the N-terminus of the intact protein.

In this way, the oligonucleotides effectively "bracket" the region of the P. vulgaris chromosome that encodes the N-terminal region of the chondroitinase I gene. It is worth noting that the PCR methodology offers an extremely large potential amplification of this bracketed region. Thirty PCR cycles, in theory, increase the number of copies of this DNA segment by a factor of one billion. This allows the use of very small quantities of P. vulgaris genomic DNA as a template which will yield, potentially, microgram amounts of synthesized product which can be readily visualized, isolated and cloned.

Using the above logic, oligonucleotide mixtures are designed based on the following amino acid sequence that is found within the "18 kD" peptide: Glu-Ala-Gln-Ala-Gly-Phe-Lys (SEQ ID NO:2, amino acids 138-144). This heptapeptide is encoded by the following nucleotide sequences:

5'-GAR-GCN-CAR-GCN-GGN-TTY-AAR-3' (SEQ ID NO:15).

The complementary strand, therefore, has the following sequences:

5'-YTT-RAA-NCC-NGC-YTG-NGC-YTC-3' which is the same as 5'-(CT)TT-(AG)AA-(GATC)CC-(GATC)GC-(CT)TG-(GATC)GC-(CT)TC-3' (SEQ ID NO:16).

Using the same criteria as described above

- 48 -

for the first set of eight oligonucleotides, a further set of eight oligonucleotides (each made up of 16 unique sequences) is designed, where the individual sets of oligonucleotides have the following sequences:

5

9. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-GGC-TTC-3'
(SEQ ID NO:17)

10. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-AGC-TTC-3'
(SEQ ID NO:18)

10

11. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-TGC-TTC-3'
(SEQ ID NO:19)

12. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-CGC-TTC-3'
(SEQ ID NO:20)

15

13. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-GGC-CTC-3'
(SEQ ID NO:21)

14. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-AGC-CTC-3'
(SEQ ID NO:22)

15. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-TGC-CTC-3'
(SEQ ID NO:23)

20

16. 5' -TT-GAA- (AG) CC- (GATC) GC- (CT) TG-CGC-CTC-3'
(SEQ ID NO:24)

25

Unlike oligonucleotides 1-8 above, one base is deleted from the 5' end of oligonucleotides 9-16 in order to reduce the number of sequence permutations.

30

In this case, one pool has a perfect match for the first eight nucleotides at the 3'-end, while 50% of this same pool has an eleven-nucleotide perfect match with the genomic DNA of *P. vulgaris* encoding chondroitinase I.

35

For a third set of oligonucleotide mixtures, the following amino acid sequence, obtained as part of the N-terminal amino acid sequence of the "90 kD" peptide, is used: Gly-Ala-Lys-Val-Asp-Ser (SEQ ID NO:2, amino acids 189-194). This hexapeptide can be

- 49 -

encoded by the following nucleotide sequences:

5'-GGN-GCN-AAR-GTN-GAY-TCN-3' (SEQ ID NO:25)

or

5'-GGN-GCN-AAR-GTN-GAY-AGY-3' (SEQ ID NO:26)

The complement of this sequence is:

5'-NGA-RTC-NAC-YTT-NGC-NCC-3' (SEQ ID NO:27)

or

5'-RCT-RTC-NAC-YTT-NGC-NCC-3' (SEQ ID NO:28)

These possible sequences are represented
using the following oligonucleotide mixtures:

17. 5'-GA-GTC-(GATC)AC-(TC)TT-(AG)GC-GCC-3' (SEQ
ID NO:29)

18. 5'-GA-GTC-(GATC)AC-(TC)TT-(AG)GC-ACC-3' (SEQ
ID NO:30)

19. 5'-GA-GTC-(GATC)AC-(TC)TT-(AG)GC-TCC-3' (SEQ
ID NO:31)

20. 5'-GA-GTC-(GATC)AC-(TC)TT-(AG)GC-CCC-3' (SEQ
ID NO:32)

21. 5'-GA-GTC-(GATC)AC-(TC)TT-(TC)GC-GCC-3' (SEQ
ID NO:33)

22. 5'-GA-GTC-(GATC)AC-(TC)TT-(TC)GC-ACC-3' (SEQ
ID NO:34)

23. 5'-GA-GTC-(GATC)AC-(TC)TT-(TC)GC-TCC-3' (SEQ
ID NO:35)

24. 5'-GA-GTC-(GATC)AC-(TC)TT-(TC)GC-CCC-3' (SEQ
ID NO:36)

Unlike oligonucleotides 1-8 above, one base
is deleted from the 5' end of oligonucleotides 17-24
in order to reduce the number of sequence

- 50 -

permutations.

In this case, one oligonucleotide mixture has half of its members perfectly matched for the first eight nucleotides at the 3'-end, and one quarter of the oligonucleotides in the pool are perfectly matched for eleven nucleotides at the 3'-end.

These twenty-four oligonucleotide mixtures are purchased from Biosynthesis, Inc. (Denton, TX), and are provided as fully deprotected, purified and lyophilized samples. In each case (except oligonucleotide #20), 5 O.D. units of synthetic DNA are obtained. This is resuspended in 0.5 ml of water to yield a solution that contains approximately 50-60 pmoles of oligonucleotide per microliter. The remaining sample (oligonucleotide #20) contains 15 O.D. and is resuspended with one ml of water to give a solution with approximately 90 pmole/ μ l.

A typical 50 μ l PCR reaction contains approximately 20 ng of *P. vulgaris* genomic DNA as template; 200 μ M each of dATP, dGTP, dCTP, dTTP; 50mM KCl; 10mM Tris-HCl (pH 8.4); 1.5 mM MgCl₂; 0.01% gelatin; 2.5 units of Ampli-Taq™ DNA polymerase (Perkin-Elmer/Cetus, Norwalk, CT); and 50 pmoles of each oligonucleotide pool to be tested. The reactions are overlaid with mineral oil (Plough) and incubated in a Perkin-Elmer/Cetus Thermalcycler™.

For each cycle, the instrument is programmed to denature the template DNA at 94°C for 1.25 minutes, anneal the oligonucleotide primers to the denatured template at 60°C or 62°C for one minute, and to extend these primers via DNA synthesis at 72°C for 2.25 minutes. Thirty such cycles are carried out in an experimental amplification. The products are analyzed by running an aliquot on a 4% NuSieve™ (FMC Biochemicals, Rockland, ME) GTG gel containing

- 51 -

approximately 0.5 μ g/ml ethidium bromide using either Tris-borate or Tris-acetate buffers at either full or half strength. These gels are usually run overnight at approximately 1V/cm and photographed on a long wavelength UV transilluminator using a red filter and Polaroid Type 57 film.

PCR experiments are run testing the pairwise combinations between oligonucleotide pools #1-8 (derived from the "110 kD" amino-terminal sequence of chondroitinase I), pools #9-16 (derived from a peptide sequence contained within the "18 kD" fragment), and pools #17-24 (derived from the amino-terminal sequence of the "90 kD" fragment). The most effective amplifications observed (based on the visual yield of a discrete DNA band detected on gel electrophoretic analysis of the reaction products) are between oligonucleotide pools #4 and #18, and pools, #4 and #9,10,11, or 12. In general, the other pools, which differ by one nucleotide from these pools, also yield some amplification. A difference of two nucleotides results, essentially, in no observed product. It is important to note, however, that the annealing temperatures are deliberately kept at 60-62°C to enhance such discrimination.

PCR amplifications using oligonucleotide pools #4 and #18 yield a product of approximately 500 bp as estimated relative to size standards (pBR322 digested with MSP-1 (New England Biolabs, Beverly, MA) ranging from 30 to 700 bp on NuSieve™ agarose gels. The product from the use of oligonucleotide pool #4 combined with pools #9, 10, 11, or 12 is approximately 350 bp in length. Furthermore, the larger product could be isolated from an agarose gel, diluted a thousand-fold, and then used as the template in a second PCR reaction employing oligonucleotide pools #4

- 52 -

and #9 as primers, which yield a product of approximately 350 bp. That is, the smaller PCR product is synthesized from the larger one in agreement with what would be expected if these sequences were all derived from the P. vulgaris chondroitinase I gene. This indicates that the desired region of the genome is amplified.

The larger PCR product is isolated from an agarose gel using a Qiaex™ extraction procedure according to the manufacturer's instructions (Qiagen, Chatsworth, CA). The isolated DNA is then subjected to a "fill-in" reaction (11) to remove the extra, protruding adenine residue that Taq DNA polymerase tends to add to the 3'-end of DNA in a template-independent reaction (12). The isolated DNA is then treated with T₄ polynucleotide kinase to add a phosphate moiety to the 5'-ends of the PCR products to allow them to be joined to the vector DNA. After these treatments, the PCR product is ligated to pIBI24, a high copy vector containing a polylinker (IBI, New Haven, CT), that is first sequentially digested with PstI, "filled-in" and then treated with calf intestinal alkaline phosphatase (Boehringer-Mannheim).

Once the PCR product is cloned into pIBI24, it is removed as an EcoRI-HindIII fragment by virtue of the restriction sites within the polylinker carried by the plasmid. This fragment is then cloned into both M13mp18 and M13mp19 (13; New England Biolabs, Beverly, MA) after cleavage with both EcoRI and HindIII and then phosphatased. Single stranded DNA corresponding to these constructions is then isolated and subjected to DNA sequence analysis using an Applied Biosystems (Foster City, CA) instrument and Taq sequencing kit. The results indicate that the

- 53 -

larger PCR product is 455 bp in length. As expected, the ends of the fragment are derived from the oligonucleotide pools used as primers.

The DNA sequence is translated into an uninterrupted amino acid sequence that is in agreement (with one exception described below) with the available data obtained by amino acid sequence analysis on the native chondroitinase I protein itself, including, for example, a twelve residue oligopeptide (SEQ ID NO:2, amino acids 133-144). An eight residue oligopeptide derived from the DNA sequence (SEQ ID NO:2, amino acids 71-78) also matches a previously sequenced oligopeptide derived by a combination of trypsin digestion and cyanogen bromide treatment of the native protein. The only discrepancy between the two sequences is at amino acid residue #162 of the mature protein (SEQ ID NO:2, amino acid 186), where the DNA sequence codes for an arginine, while the native protein sequence indicates a leucine.

Since a single nucleotide alteration would change a leucine codon (CTT) to an arginine codon (CGT), an initial interpretation suggests that this may be caused by a lack of perfect incorporation fidelity by the Taq DNA polymerase during the in vitro amplification process. However (see below) later results indicate that the DNA sequence is correct, rather than the amino acid sequence obtained by analyzing the native enzyme. These results also indicate that the "18 kD" and the "90 kD" fragments are, in fact, contiguous pieces of the chondroitinase I protein that has been cleaved (presumably by a contaminating protease), predominately between residues #157 (Gln) and #158 (Asp) of the mature protein (SEQ ID NO:2, between amino acids 181 and 182). All of the above information supports the

- 54 -

interpretation that the cloned DNA (at least that portion that is bracketed by the oligonucleotide primers) generated by PCR amplification represents part of the authentic chondroitinase I gene of P. vulgaris and, therefore, can be used as a probe to identify cosmid clones that carry the intact gene.

Although it is possible to isolate the entire gene coding for a protein of interest using PCR amplification (thereby avoiding construction of a gene bank and many of the other steps described below) by employing oligonucleotide primers derived from the amino-terminus of the protein coupled with primers derived from the carboxyl-terminal amino acid sequence, there are several potential problems in this approach. In the case of the P. vulgaris chondroitinase I, the problems include: (1) the assumption that the protein being sequenced has not been processed at either end (not likely to be true, for example, with a secreted protein), (2) the occasional lack of fidelity exhibited by Taq DNA polymerase during PCR reactions, and (3) the rather large size of the bracketed region of the DNA that is to be amplified which was expected to be approximately 3000 bp (deduced from the apparent molecular weight of approximately 110 kD). Consequently, the approach of constructing a gene bank is selected.

Example 3

Generation Of A Labeled Probe, Colony Hybridization And Identification Of Positive Cosmid Clones From The P. vulgaris Gene Bank

The cloned PCR product corresponding to the 455 bp near the amino-terminal coding portion of the P. vulgaris chondroitinase I gene is released from the

- 55 -

plasmid DNA into which it had been cloned by digestion with the restriction enzyme SalI. This is a consequence of the presence of one SalI site within the polylinker sequence and a second SalI site within the cloned PCR amplification product (this is fortuitous in that the latter SalI site is derived from the nucleotide sequence of oligonucleotide pool #18 near its 5'-end; in fact, there is no recognition site for SalI within the P. vulgaris chondroitinase I gene itself). A total of approximately 260 µg of plasmid DNA is digested with SalI and the products separated by electrophoresis on a NuSieve™ GTG agarose gel. The desired approximately 450 bp fragment is isolated using a Qiaex™ extraction protocol. The fragment is then denatured by heating at 95-100°C for 5-15 minutes, followed by rapid cooling. The denatured fragment is then labelled with digoxigenin-labelled dUTP (Boehringer-Mannheim, Indianapolis, IN) in two 200 µl reactions.

Aliquots of the six P. vulgaris cosmid gene banks described in Example 1 above are used to infect the E. coli strain ER1562 described above and a total of approximately 10,000 colonies are obtained on the appropriate selective plates. These colonies (on a total of 50 plates) are replica plated onto two nylon membranes on selective agar as well as to a third selective plate. After overnight incubation, the colonies on the filters are lysed by sequentially treating with 10% sodium dodecyl sulfate (SDS) and 0.5 M NaOH for 5-30 minutes each. The cells from the lysed colonies are neutralized by being placed on sheets saturated with 1 M Tris-HCl (pH 7.4) (twice) and then on paper saturated with 2X standard saline citrate prior to vacuum drying at 80°C. The DNA from the lysed colonies is then fixed to the membranes.

- 56 -

The filters are then washed by incubation of the filters at 42°C with agitation for 1-3 hours, using at least 10 ml/filter of 0.05 M Tris HCl, 0.5-1 M NaCl/0.001 M EDTA, pH 8, 0.1% SDS and 0.05 mg/ml proteinase K. The filters are then rinsed with 2 X SSC and pre-hybridized by incubation with a hybridization buffer at 65°C for 1-3 hours. The filters are then hybridized overnight at 65-68°C using the digoxigenin-labeled probe described above (0.5-50 ng/ml in a hybridization solution). The hybridized filters are washed with SSC and SDS, re-blocked with a blocking reagent (Component #11 of DNA Labelling and Detection Kit, Nonradioactive, Boehringer Mannheim, Indianapolis, IN) and exposed to polyclonal sheep anti-digoxigenin Fab fragments conjugated to alkaline phosphatase.

The positive clones are visualized by incubation of the antibody-labeled filters in the presence of BCIP (bromo-chloro-indolyl-phosphate) and NBT (nitro-blue tetrazolium). The presence of the desired DNA fragment within a colony will result in a dark brownish-purple spot in the filter after this hybridization procedure. After approximately four hours, the developed filters are used as templates to guide the selection of a total of 117 clones which are then picked to selective media. A small-volume (10 ml) culture ("Miniprep") of each of these clones is grown in selective media and plasmid DNA is then isolated using materials and protocols supplied by Qiagen.

- 57 -

Example 4

Restriction Mapping And Southern Hybridization
Used To Localize The Position Of The
Chondroitinase I Gene Within Individual Clones

5

A number of approaches are used to guide the selection of particular cosmid clones for further study. One is to carry out Southern hybridization (8) using the same PCR-generated fragment as a probe against P. vulgaris genomic DNA that had been digested by a number of restriction enzymes and then fractionated on an agarose gel prior to transfer to a nylon membrane. The probe is labeled with digoxigenin-dUTP by including this nucleotide analogue in a PCR amplification. In this reaction, the gel-purified product of a previous PCR amplification (that using P. vulgaris genomic DNA as template) is diluted 10,000-fold and serves as the template in a second PCR amplification.

20

This latter reaction is made up as a 0.5 ml mixture, which is then divided into ten individual tubes and amplified as described above for 25 cycles using oligonucleotide pools #2 and #10 (see above) as the primers. The normal complement of deoxyribonucleoside triphosphates is replaced with a digoxigenin-dUTP labeling mixture from the manufacturer (Boehringer-Mannheim, Indianapolis, IN), which yields a final concentration of 100 μ M each of dATP, dCTP and dGTP, 65 μ M dTTP and 35 μ l digoxigenin-dUTP. The reactions are pooled and precipitated according to the manufacturer's recommendations. An aliquot of the resuspended product is examined by gel electrophoresis and exhibits a single band between approximately 300 and approximately 400 bp in length as expected for the "smaller" PCR product described

35

- 58 -

above.

To avoid problems encountered with the highly viscous *P. vulgaris* genomic DNA preparation, the DNA (approximately 5 μ l) is diluted into large (0.35 ml) volumes for digestion with the various restriction enzymes. The DNA is then concentrated by ethanol precipitation prior to fractionation on agarose gels and transfer to nylon membranes. The data obtained in these experiments indicates that the chondroitinase I gene (at least that portion that hybridizes to the N-terminal coding region represented by the probe described above) is carried on a BstYI fragment of approximately 2800 bp, an EcoRV fragment of 5400 bp, and on large (equal to or greater than approximately 10kb) DNA fragments generated by NsiI, BglII, HindIII, and StyI.

Large scale cultures (500 ml) of a number of hybridizing cosmid clones are grown and plasmid DNA is isolated from these cultures for use in mapping the location of the chondroitinase I gene. The DNA of the gene is expected to represent only approximately 10 per cent of the *P. vulgaris* DNA carried within each cosmid. A number of these clones are digested with BstYI and NsiI and the products are fractionated on an agarose gel. Individual fragments are then isolated, a portion tested for the presence of chondroitinase I sequences by Southern hybridization, and then subcloned into appropriate vectors.

Two of these fragments are of special interest. The first, a BstYI fragment of approximately 2800 bp, is observed in a number of cosmid clones, including those designated #2 and #45. The DNA isolated from these two cosmid clones is designated LP² 751 and LP² 760. With LP² 760, the approximately 2800 bp BstYI fragment is well separated

- 59 -

from the other BstYI fragments and is therefore more readily subcloned into another vector designated pT660-3. The plasmid designated pT660-3 is a derivative of pBR322 in which the DNA from a point immediately downstream of the promoter for tetracycline resistance (approximately bp 80) as far as the PvuII site (approximately bp 2070) is deleted and replaced with a BamHI linker. Similarly, the approximately 10 kb NsiI fragment (which hybridizes with the chondroitinase probe described above) is readily isolated from a digest performed on LP² 751. These two fragments are referred to as the "2800 bp BstYI" fragment and the "10 kb NsiI" fragment.

The 2800 bp BstYI fragment is small enough to permit a second restriction enzyme digestion on this piece of DNA in order to obtain a fragment suitable for DNA sequence analysis. This is important because the hybridization experiments serve to identify the N-terminal coding region of the chondroitinase I gene, due to how the probe is derived. This procedure does not, however, indicate to which side the rest of the gene is located. Given the relative size of the probe (less than 500 bp) compared to the predicted size of the intact gene (greater than 3000 bp), this is not a trivial consideration. The nucleotide sequence, however, clearly indicates in which direction the gene would be "read" and therefore, which restriction fragments should be cloned in order to obtain the entire gene.

The subcloned 2800 bp BstYI fragment contains two internal EcoRV sites, which suggests that the resulting fragments might be small enough for DNA sequencing. However, the EcoRV sites are symmetrically placed within the 2800 bp BstYI fragment; each EcoRV site is approximately 1200 bp

- 60 -

from one end, with the space between them equal to approximately 400 bp. The subcloned fragment is digested asymmetrically by taking advantage of unique restriction sites present within the vector. In this manner, the "halves" of the 2800 bp BstYI fragment are distinguished physically and, by Southern hybridization, the "end" that contains the chondroitinase I N-terminal coding region is ascertained. Once this is done, the appropriate piece, which is a HindIII-EcoRV fragment of approximately 1200 bp, is subcloned into both M13mp18 and M13mp19 vectors which are first digested with both HindIII and SmaI and subsequently treated with calf intestinal alkaline phosphatase. The DNA sequence derived from these subclones reveals a number of features that clearly establish the location of the chondroitinase I gene, as well as the direction in which it is read.

Starting with nucleotide #183 in this sequence (SEQ ID NO:1, nucleotide 191), a coding region is observed which matches the first thirty previously-identified amino acids of the P. vulgaris chondroitinase I enzyme. Preceding this sequence, it is possible to discern a number of other features by their analogy to corresponding sequence motifs from previously analyzed E. coli genes. These features include: (1) nucleotides 32-37 (SEQ ID NO:1, nucleotides 40-45) which match in three of six positions with the consensus "-35" region of a promoter and, after a 17 nucleotide space, a "-10" region of a promoter (matching in six of seven positions with the consensus "-10" region); (2) a putative "Shine-Dalgarno" sequence can be noted between nucleotides 98-103 (SEQ ID NO:1, nucleotides 106-111); and (3) there is an in-frame ATG initiation codon at nucleotides 111-113 (SEQ ID NO:1, nucleotides

- 61 -

119-121), which indicates that the P. vulgaris chondroitinase I enzyme is synthesized with a 24 amino acid signal sequence which is, presumably, removed as the protein is transported across the inner membrane.

5 The second fragment that is subcloned (into a pIBI24 derivative that is first modified to include an NsiI restriction site in place of the PstI site normally present in the polylinker of this vector) is the approximately 10 kb NsiI fragment. Digestion of
10 this approximately 14 kb recombinant molecule (the approximately 10 kb NsiI fragment in pIBI24) with EcoRV yields four fragments of approximately 9 kb, 2.3 kb, 2.1 kb, and 0.4 kb. Southern hybridization
15 analysis using the probe derived from the N-terminal amino acid sequence indicates that the related chondroitinase gene sequences are contained within the largest fragment (the approximately 9 kb EcoRV fragment).

 Since there is no other fragment larger than
20 2.9 kb (the size of pIBI24 which has no internal EcoRV recognition sites), this approximately 9 kb EcoRV fragment must contain the vector as well as P. vulgaris DNA. A double digestion of this recombinant molecule with NsiI and EcoRV releases the pIBI24
25 vector as a 2.9 kb fragment; it also yields fragments of approximately 4.5 kb, 2.3 kb, 2.1 kb, 1.0 kb and 0.4 kb. Taken together (along with the information presented above on the 2.8 kb BstYI fragment which has two internal EcoRV sites separated by approximately
30 0.4 kb), an initial restriction map is constructed.

 A double digestion with EcoRV and HindIII releases fragments of approximately 4.1 kb, 2.3 kb, 2.1 kb, 2.0 kb, 1.3 kb, 1.1 kb and 0.4 kb. Three of these fragments (2.3 kb, 2.1 kb, and 0.4 kb) are
35 apparently EcoRV fragments that have not been cut by

- 62 -

HindIII. Again, the only fragment larger than the vector (4.1 kb) indicates that this fragment includes pIBI24 (2.9 kb). The approximately 2.0 kb fragment hybridizes with the chondroitinase probe, thereby serving to place one of the HindIII sites. Since there is a HindIII site in the polylinker, it too can be placed, leaving the last HindIII site to be placed by deduction.

Double digestion of the cloned approximately 10 kb NsiI fragment with EcoRV and EcoRI yields six fragments (of approximately 4.2 kb, 3.5 kb, 2.3 kb, 2.1 kb, 1 kb, and 0.4 kb), indicating the presence of two EcoRI sites -- one in the polylinker and one in the cloned *P. vulgaris* DNA. Southern hybridization reveals that the approximately 4.2 kb band in this double digest contains the chondroitinase I N-terminal coding sequence. Adding this information to the above data yields a preliminary restriction map for the subcloned approximately 10 kb NsiI fragment in pIBI24 (Figure 1).

It should be noted that, in further support of the placement and orientation of the chondroitinase I gene, *in vitro* chondroitinase I assays in which the activity of the enzyme based on measuring the release of unsaturated disaccharide from chondroitin sulfate C at 232 nm are carried out on a small number of samples. In one case, an aliquot of an overnight culture used to prepare LP² 751 (ER1562 carrying cosmid DNA selected from the colony hybridizations) is found to express 0.12 units/ml of chondroitinase. In addition, one of the EcoRV-deletion constructions (to be described below) is grown overnight in the presence of ampicillin. This culture is then inoculated into fresh selective media either with or without isopropyl-beta-D-thiogalactopyranoside (IPTG) which is

- 63 -

expected to increase the level of transcription from the lac promoter present in pIBI24. The assay results of 0.29 units/ml of chondroitinase without and 0.36 units/ml with IPTG induction indicate that, even after the EcoRV deletion, the gene is still intact and possibly oriented in the same direction as that of the lac promoter.

Although the sizes of the fragments in the above discussion are approximate (especially the approximately 1 kb region between the EcoRI/NsiI in the polylinker and the nearest EcoRV site; in addition, there also might be another small EcoRV fragment that is still unmapped), overall they suggest that the approximately 4.2 kb EcoRV-EcoRI fragment contains the entire chondroitinase I gene. In order to facilitate the restriction mapping, an EcoRV deletion is constructed using the approximately 10 kb NsiI fragment cloned into pIBI24 (LP²776). This DNA is digested with EcoRV, treated with calf intestinal alkaline phosphatase, and fractionated on an agarose gel. The largest (approximately 10kb) fragment is extracted from the gel and ligated together in the presence of a phosphorylated EcoRI linker. The resulting construction (LP² 786) is next digested with EcoRI to yield three fragments. Although it is not completely separated from the pIBI24-containing, somewhat smaller fragment, an approximately 95% homogenous, approximately 4.2 kb EcoRI fragment is obtained after extraction from the gel. This EcoRI fragment is then used for DNA sequence analysis.

- 64 -

Example 5DNA Sequence Analysis Of The Approximately
4.2 kb EcoRI Fragment

5 The approximately 10 kb NsiI fragment,
cloned into pIBI24, is digested with EcoRV (as
described above) and ligated together in the presence
of EcoRI linkers. The net result of this construction
is the deletion of approximately 5 kb of P. vulgaris
10 DNA from this subcloned piece of DNA and the
simultaneous introduction of another EcoRI site into
the molecule. One hundred micrograms of this "EcoRV
deletion" construction (LP² 786) is digested with EcoRI
and fractionated on an agarose gel. The desired
15 approximately 4.2 kb fragment is eluted from the gel,
precipitated and resuspended in 150 μ l TE described
above. One-third of this material is then ligated to
itself (polymerized) and, after destruction of the DNA
ligase by heating, the DNA is sonicated to generate
20 random, small pieces suited to DNA sequence analysis.

The ends are rendered flush in a "fill-in"
reaction mediated by the "Klenow fragment" (10; New
England Biolabs, Beverly, MA) of the DNA polymerase I
of E. coli, and then ligated into SmaI-cut and
25 phosphatased M13mp19. This recombinant DNA is used to
transform the male E. coli strain MV1190 and 500 of
the phage plaques obtained are picked into SM buffer
(NaCl, 100 mM, MgSO₄, 8 mM, Tris-HCl, pH 7.4, 50 mM and
0.01% gelatin) to serve as stocks for the infection of
30 small (less than or equal to 10 ml) cultures that are
then used for the isolation of single stranded
template DNA.

DNA sequencing is carried out at elevated
temperatures using Taq DNA polymerase and
35 fluorescently-labeled oligonucleotide primers. The

- 65 -

data are collected using a Model 370A DNA sequencing system (Applied Biosystems, Foster City, CA).

Sequence editing, overlap determinations and derivation of a consensus sequence are performed using a collection of computer programs obtained from the Genetics Computer Group at the University of Wisconsin (14). The resulting DNA sequence of this EcoRI fragment is 3980 nucleotides in length (SEQ ID NO:1). It is to be noted that the EcoRI site near the N-terminal coding sequence is derived from the linker ligated into this site; it is not present in the P. vulgaris chromosome. This position actually is an EcoRV site in the cloned cosmid DNA.

Translation of the DNA sequence into the putative amino acid sequence reveals a continuous open reading frame encoding of 1021 amino acids (SEQ ID NO:2), with a 24 residue signal sequence (SEQ ID NO:2, amino acids 1-24), followed by a 997 residue coding sequence for the mature (processed) chondroitinase I protein (SEQ ID NO:2, amino acids 25-1021). Computer analysis using the programs described above of this sequence predicts a molecular weight of 115,090.94 for the unprocessed protein, a molecular weight of 112,507.82 for the mature "110 kD" (transported) protein, 17,503.43 for the first 157 amino acids (the "18 kD" fragment) (SEQ ID NO:2, amino acids 25-181) and 95,022.40 for the remaining 840 amino acids (the "90 kD" fragment) (SEQ ID NO:2, amino acids 182-1021) and a molecular weight of 2601.14 for the 24-residue signal sequence. One notable feature of the amino acid composition is the absence of cysteine which could be important if the protein has to be re-folded at any point.

In the nucleotide sequence, it was noted above that there is a unique SphI restriction site

- 66 -

located approximately 230 bp beyond the end of the gene (SEQ ID NO:1, nucleotides 3414-3419), which presents a unique target site that can be manipulated to allow the facile movement of the gene to achieve the overall goal of expressing chondroitinase at high levels in E. coli. Although there are two recognition sites for ClaI (ATCGAT), one of them (SEQ ID NO:1, nucleotides 2702-2707) is embedded within the E. coli dam recognition sequence (GATC) (SEQ ID NO:1, nucleotides 2701-2704). The resulting adenine methylation by the dam-encoded enzyme blocks cleavage of this site by ClaI; therefore, there is, in effect, a "unique" ClaI site (SEQ ID NO:1, nucleotides 497-502) which is used, as described below, to reconstruct the chondroitinase I gene after the appropriate site-specific mutageneses are carried out.

Example 6

Site-specific Mutagenesis Of The Cloned P. vulgaris Chondroitinase I Gene

The site-specific mutagenesis method employed is based on that of Kunkel (15), using materials purchased from Bio-Rad, Melville, N.Y. (Muta-Gene™ In Vitro Mutagenesis Kit). In this procedure, the target DNA to be mutagenized is first cloned into an appropriate M13-derived vector. In this case, the recombinant molecule used (M13mp19 into which is cloned the approximately 1200 bp EcoRV-HindIII fragment as described above) encompasses the N-terminal coding region of the chondroitinase I gene. This recombinant phage is replicated in the E. coli host strain CJ236 (Bio-Rad), a male strain that carries the dut and ung alleles. The combination of these two mutations, dut (dUTPase) and ung (uracil-N-

- 67 -

glycosylase), results in the incorporation of some uracil, rather than thymine, residues into the DNA synthesized in this organism. Single stranded template is then isolated after propagation on CJ236 and an appropriate, mutagenic, synthetic oligonucleotide is annealed to this DNA.

This oligonucleotide serves as a primer for T7 DNA polymerase which copies the entire recombinant molecule. T4 DNA ligase is then used to seal the nick between the first residue of the mutagenic oligonucleotide and the last residue added in vitro. The newly synthesized DNA (containing the desired base changes) therefore does not contain uracil, while the template DNA does. Transformation of a non-mutant (with respect to the dut and ung alleles) male E. coli strain yields phage progeny that are primarily derived from the mutagenized strand synthesized in vitro as a result of the inactivation of the uracil-containing template strand.

In this specific case, four resuspended plaques (aliquots of which had been used for DNA sequencing which established the N-terminal coding region of the chondroitinase I gene and included another 110 bp "upstream" of the presumed translation initiation site (see above)) are used to infect the male host strain CJ236 (dut ung). Individual plaques are picked to 0.5 ml of phage dilution buffer (PDB). One picked plaque from each transformation is adsorbed to log phase CJ236 and the infected culture grown for 6.5 hours. The cells are pelleted by centrifugation, and the supernatant heated to 55°C for 30 minutes and then stored at 4°C. Single stranded DNA is isolated from 100 ml of each supernatant and resuspended in a total volume of 0.1 ml of TE.

The goal of the site-specific mutagenesis is

- 68 -

to modify the "ends" of this gene to allow it to be moved, precisely, into an appropriate high-level E. coli expression system. The target vector chosen (pET9-A; see above) is one derived from genetic regulatory elements present in the bacteriophage T7. In this system, there is a unique NdeI site (CATATG) that includes the translation initiation codon as well as a downstream BamHI site that, together, allow the direct, unidirectional, insertion of a gene encoding the protein that is to be expressed. These two sites are preceded by a T7-specific promoter sequence and trailed by a transcription terminator that functions with the T7 RNA polymerase. Accordingly, these two restriction sites (NdeI and BamHI) are introduced into the cloned gene for P. vulgaris chondroitinase I.

In order to introduce the NdeI site (containing the ATG initiation codon) both before the signal sequence as well as, in a second construction, before the coding sequence for the mature protein (thereby deleting the signal sequence), two synthetic oligonucleotides are designed and synthesized (purchased from BioSynthesis, Inc., Denton, TX). The first, designated oligonucleotide # 25 (SEQ ID NO:37), retains the signal sequence while the second, oligonucleotide #26 (SEQ ID NO:38), deletes the signal sequence and allows the direct expression of the mature chondroitinase I protein (which can have an additional methionine residue at the N-terminus (SEQ ID NO:5, amino acid number 1)).

The native sequence, including the predicted initiation codon, is presented on line 1 below while the mutagenic oligonucleotide #25 (which differs in the three nucleotides immediately upstream of the initiation codon) is presented on line 2:

- 69 -

1) 5'-GCCAGCGTTTCTAAGGAGAAAAATAATGCCGATATT-
TCGTTTTACTGC-3' (SEQ ID NO:1, nucleotides
94-141)

5 2) 5'-GCCAGCGTTTCTAAGGAGAAAACATATGCCGATATT-
TCGTTTTACTGC-3' (SEQ ID NO:37)

10 For the construction in which the signal
sequence is deleted, the site-specific mutagenesis is
carried out at the junction of the signal sequence and
the start of the mature protein (line 3) using the
mutagenic oligonucleotide # 26 (line 4) (which differs
by six nucleotides, including the location of the
initiation codon):

15 3) 5'-GCGCCTTATAACGCGATGGCAGCCACCAGCAATCCTG-3'
(SEQ ID NO:1, nucleotides 170-206)

20 4) 5'-GCGCCTTATAACGCGCATATGGCCACCAGCAATCCTG-3'
(SEQ ID NO:38)

25 The underlined GCC in line 3 corresponds to
the codon for alanine which is the N-terminal amino
acid for the mature, processed form of the *P. vulgaris*
chondroitinase I.

30 In order for these oligonucleotides to be
used, their 5'-ends need to be phosphorylated. There-
fore, oligonucleotide # 25 (5 O.D. units) is
resuspended with 0.5 ml of TE, while oligonucleotide #
26 (also 5 O.D. units) is resuspended in 0.65 ml TE to
yield stocks that are approximately 20 nM, i.e., 20
pmole/ μ l. Three nanomoles (150 μ l of stock solution)
of each oligonucleotide are kinased in separate (0.35
35 ml) reactions containing 35 μ l 10x ligase salts (New
England Biolabs, Beverly, MA): 0.5 M Tris-HCl (pH

- 70 -

7.8), 0.1 M MgCl_2 , 0.2 M dithiothreitol, 10 mM ATP, 0.5 mg/ml bovine serum albumin), 35 μl 0.1 M dithiothreitol, 10 μl (100 units) T4 polynucleotide kinase (New England Biolabs) and made up to volume with 120 μl TE. The reactions are incubated at 37°C for 40 minutes and the enzyme inactivated at 70°C for 20 minutes.

Template DNA (5 μl of the preparation described above) and phosphorylated mutagenic primer (approximately 2 pmole) are annealed in a 20 μl volume containing 20 mM Tris-HCl (pH 7.4), 2 mM MgCl_2 , and 50 mM NaCl. The sample is heated at 70°C for 45 minutes in a Perkin-Elmer/Cetus Thermalcycler™. The sample is then gradually cooled from 70°C to 25°C over a 45 minute period. The annealed mixture is placed on ice and the following components added: 2 μl of 10 X synthesis buffer (Bio-Rad): 5mM each of dATP, dGTP, dCTP, dTTP; 10 mM ATP; 100mM Tris-HCl (pH 7.4); 50 mM MgCl_2 ; 20 mM dithiothreitol), 2 μl of T4 DNA ligase (6 units) and 1 μl of T7 DNA polymerase (1 unit). These reactions are incubated for 5 minutes each at 0°C (on ice), 11°C, 25°C, and finally for 30 minutes at 37°C. The reactions are stopped by the addition of 75 μl of 10 mM Tris-HCl-10 mM EDTA (pH 8.0) and placed at -20°C.

After the mutagenized DNA is thawed, it is used to transform the male *E. coli* strain MV1190 (*dut⁺ ung⁺*). Individual plaques obtained are picked and single-stranded DNA is isolated and sequenced. For those cases in which the desired sequence changes are introduced, another aliquot of the resuspended plaque is used to infect strain MV1190, but in this case the intracellular, double-stranded replicative form of the recombinant DNA is isolated from the infected cell pellets using the Mini-Prep procedure referenced above.

- 71 -

Example 7

Reconstruction Of The Site-Specifically
Mutagenized Chondroitinase I Gene And Its
High-Level Expression In E. coli

5

Example 6 described the site-specific mutageneses that created an NdeI site immediately preceeding the signal sequence, as well as a second construction which placed the NdeI site adjacent to the triplet which codes for the N-terminal alanine found on the mature, processed P. vulgaris chondroitinase I gene. In each case, the ATG sequence of the NdeI recognition site (CATATG) can function as the translation initiation codon for the protein (either with or without the signal sequence).

15

In order to transfer these alterations from the M13 vector in which they were constructed, to the full chondroitinase I gene, the isolated replicative form is digested with KpnI and ClaI. The KpnI site is part of the M13mp19 polylinker, while the ClaI site is found approximately 490 bp from the end of the cloned fragment of the chondroitinase I gene. The restriction digestion products obtained are fractionated on a 4% NuSieve™ GTG agarose gel run in 1/2 X Tris-Acetate buffer (TAE). The appropriate approximately 500 bp band is extracted from the gel using Qiaex™. Similarly, plasmid DNA (LP² 786) carrying the chondroitinase I gene is also digested with KpnI and ClaI and then fractionated on a 0.8% agarose gel run in 1/2 X TAE. In this case, the KpnI site is part of the polylinker of pIBI24, while the ClaI site corresponds to the one described above. (As stated above, there is a second ClaI site in the chondroitinase I gene, but it is not cleaved by ClaI because this site is apparently blocked by dam

20

25

30

35

- 72 -

methylation. The site-specific mutagenesis and reconstruction of the chondroitinase I gene were carried out before the entire nucleotide sequence was ascertained).

5 The approximately 7 kb fragment containing the pIBI24 vector and the large fragment of the chondroitinase I gene are isolated from the agarose gel by electroelution (11), followed by ethanol precipitation. This 7 kb fragment is then treated
10 with calf intestinal alkaline phosphatase, extracted first with phenol-chloroform, then with chloroform, and then precipitated twice with ethanol and finally resuspended with 0.1 ml TE. The two isolated N-terminal encoding fragments (the two approximately 500
15 bp KpnI-ClaI pieces containing the two site-specifically mutagenized sequences, one with and one without the signal sequence) are each ligated to the approximately 7 kb fragment encompassing the remainder of the chondroitinase I gene and the pIBI24 vector.
20 The ligase reaction is then used to transform the E. coli strain 294 and ampicillin resistant derivatives obtained. DNA is isolated from small (10 ml) cultures and digested with NdeI to verify the presence of this restriction site within the reconstructed DNA.

25 In order to remove the (apparent) P. vulgaris promoter and ribosome binding site, the modified chondroitinase I genes are isolated as approximately 4.5 kb NdeI-NsiI fragments and subcloned into a pBR322 variant in which the EcoRI site is first
30 filled-in, then dephosphorylated, and finally a phosphorylated NsiI linker (New England Biolabs) inserted. The sequence of the linker used (TGCATGCATGCA) to place the NsiI site (ATGCAT) into pBR322 also includes an SphI site (GCATGC). In order
35 to trim extra, non-coding DNA from the subcloned NdeI-

- 73 -

NsiI fragments, as well as to introduce a unique restriction site to be used later, plasmids (representing two clones each with the signal sequence retained [LP² 861 and LP² 863] and two with the signal sequence deleted [LP² 865 and LP² 867]) containing the approximately 4500 bp NdeI-NsiI segments including the chondroitinase I gene are first digested with SphI, the ends "filled-in" with the "Klenow" fragment (11) of the E. coli DNA polymerase I and the resulting DNA fragments fractionated on an agarose gel (0.8% in 1/2 X TAE). The appropriate bands (approximately 5200 bp) are eluted from the gel using Qiaex™ and then treated with calf alkaline phosphatase. After the removal of this enzyme by phenol-chloroform and chloroform extractions, the DNA is precipitated twice and finally resuspended with 0.1 ml TE.

This DNA is then ligated in the presence of a phosphorylated BamHI linker and the mixture used to transform the E. coli strain 294. Six representative, ampicillin resistant colonies from each of the four constructions are grown in small (10 ml) cultures and plasmid DNA is isolated. Digestion of the DNA from the 24 clones examined with the enzymes NdeI and BamHI indicates which contain the BamHI site and, simultaneously, releases the approximately 3400 bp NdeI-BamHI fragment which contains the chondroitinase I gene. Seventeen clones (eight with and nine without the signal sequence) yield the desired fragment which is extracted from the agarose gel with Qiaex™.

These approximately 3.4 kb NdeI-Bam-HI chondroitinase I gene-containing fragments (both with and without the signal sequence) are then used to construct a high-level expression system. The expression vector used, pET-9A (9; Novagen), is derived from elements of the E. coli bacteriophage T7.

- 74 -

It contains an origin of replication derived from the Col E1 plasmid, a kanamycin resistance determinant, and the transcription and translation initiation determinants of the T7 gene 10. The naturally-
5 occurring translation initiation codon for this gene is part of an NdeI site. This region is followed by a unique BamHI site and a T7 transcription terminator. A sample of this expression vector is digested with the restriction enzymes NdeI and BamHI,
10 dephosphorylated with calf intestinal alkaline phosphatase, and purified by agarose gel electrophoresis. Each of the chondroitinase I gene fragments (both with and without the signal sequence) is ligated to the expression vector fragment. The
15 resulting recombinant DNA mixture is used to transform the E. coli K-12 host, HMS174 (Novagen). Kanamycin-resistant colonies obtained are grown in small scale (10 ml) and plasmid DNA is isolated and examined to confirm the predicted structure.

20 Samples of these constructions are then used to transform the expression host BL21(DE3)/pLysS (10). This E. coli B strain carries the T7 RNA polymerase gene under lac control (and is therefore inducible by either lactose or IPTG) on a lambda phage integrated
25 within the E. coli chromosome, as well as the Col E1-compatible plasmid pLysS. This latter replicon specifies resistance to chloramphenicol and contains the T7 lysozyme gene inserted into the tetracycline-resistance determinant of pACYC184 (ATCC 37033,
30 American Type Culture Collection, Rockville, MD) in the "silent" orientation (read in the opposite direction relative to the tetracycline resistance gene). The T7 lysozyme is expressed at a relatively low level in this construction and serves as an
35 inhibitor of the T7 RNA Polymerase (16), thereby

- 75 -

minimizing the basal-level expression of the gene to be overexpressed.

Derivatives of BL21(DE3)/pLysS carrying the chondroitinase I gene (with the signal sequence retained and which have been subjected to the site-directed mutagenesis described in Example 6 (SEQ ID NO:3)) in pET9-A are designated LL2084, LL2085, LL2086 and LL2087. They are not tested for expression of the chondroitinase I enzyme. The native chondroitinase I gene (with the signal sequence retained) (SEQ ID NO:1), which has not been subjected to site-directed mutagenesis, is inserted into a different expression host. Expression of the chondroitinase I enzyme is achieved.

One of the derivatives of BL21(DE3)/pLysS carrying the signal-less chondroitinase I gene which has been subjected to the site-directed mutagenesis described in Example 6 (SEQ ID NO:4) inserted into pET9-A, is designated LL2088, tested and used to establish a master cell bank. The insertion of the gene into pET9-A yields the plasmid designated pTM49-6. Samples of the *E. coli* B strain BL21(DE3)/pLysS carrying the plasmid pTM49-6 constitute the deposited strain ATCC 69234.

An overnight culture of this deposited strain is grown at 30°C in the presence of 40 µg/ml of kanamycin and 25 µg/ml of chloramphenicol. A 0.5 ml aliquot of this culture is used to inoculate 100 ml of a rich "expression" medium containing M9 salts (17) supplemented with 20 g/l tryptone, 10 g/l yeast extract, and 10 g/l dextrose in addition to the same level of kanamycin and chloramphenicol.

The culture is grown at 30°C to an appropriate density (a value of 1 at A₆₀₀) and then chondroitinase I expression is induced by the addition

- 76 -

of IPTG to a final concentration of 1 mM. After three hours, samples are taken, centrifuged, and the cell pellets frozen on dry ice prior to assay. The frozen pellets are thawed, resuspended in buffer and sonicated. A value of 56 units/ml is obtained (relative to the original culture volume), which indicates that this expression system is functional. A subsequent 10 liter fermentation under controlled conditions at a higher cell density yields a maximum value of approximately 600 units/ml of chondroitinase I. This represents a substantial improvement over fermentation of the original native *P. vulgaris*, which had not expressed chondroitinase I at a level above 2 units/ml.

Example 8

Method For The Isolation And Purification Of The Native Chondroitinase I Enzyme As Adapted To The Recombinant Enzyme

The native enzyme is produced by fermentation of a culture of *P. vulgaris*. The bacterial cells are first recovered from the medium and resuspended in buffer. The cell suspension is then homogenized to lyse the bacterial cells. Then a charged particulate such as 50 ppm Bioacryl (Toso Haas, Philadelphia, PA), is added to remove DNA, aggregates and debris from the homogenization step. Next, the solution is brought to 40% saturation of ammonium sulfate to precipitate out undesired proteins. The chondroitinase I remains in solution.

The solution is then filtered using a 0.22 micron SP240 filter (Amicon, Beverly, MA), and the retentate is washed using nine volumes of 40% ammonium sulfate solution to recover most of the enzyme. The

- 77 -

filtrate is concentrated and subjected to diafiltration with a sodium phosphate buffer using a 30 kD filter to remove salts and small molecules.

The filtrate containing chondroitinase I is subjected to cation exchange chromatography using a Cellufine™ cellulose sulfate column (Chisso Corporation, distributed by Amicon). At pH 7.2, 20 mM sodium phosphate, more than 98% of the chondroitinase I binds to the column. The native chondroitinase I is then eluted from the column using a 0 to 250 mM sodium chloride gradient, in 20 mM sodium phosphate buffer.

The eluted enzyme is then subjected to additional chromatography steps, such as anion exchange and hydrophobic interaction column chromatography. As a result of all of these procedures, chondroitinase I is obtained at a purity of 90-97% as measured by SDS-PAGE scanning (see above). However, the yield of the native protein is only 25-35%, determined as described above. This method also results in the cleavage of the approximately 110 kD chondroitinase I protein into a 90 kD and an 18 kD fragment. Nonetheless, the two fragments remain non-ionically bound and exhibit chondroitinase I activity.

When this procedure is repeated with lysed host cells carrying a recombinant plasmid encoding chondroitinase I, significantly poorer results are obtained. Less than 10% of the chondroitinase I binds to the cation exchange column at standard stringent conditions of pH 7.2, 20 mM sodium phosphate.

Under less stringent binding conditions of pH 6.8 and 5 mM phosphate, an improvement of binding with one batch of material to 60-90% is observed. However, elution of the recombinant protein with the NaCl gradient gives a broad activity peak, rather than

- 78 -

a sharp peak (see Figure 2). This indicates the product is heterogeneous. Furthermore, in subsequent fermentation batches, the recombinant enzyme binds poorly (1-40%), even using the less stringent binding conditions. Batches that bind poorly are not completely processed, so their overall recovery is not quantified.

Example 9

First Method For The Isolation And Purification Of Recombinant Chondroitinase I According To This Invention

As a first step, the host cells which express the recombinant chondroitinase I enzyme are homogenized to lyse the cells. This releases the enzyme into the supernatant.

In one embodiment of this invention, the supernatant is first subjected to diafiltration to remove salts and other small molecules. An example of a suitable filter is a spiral wound 30 kD filter made by Amicon (Beverly, MA). However, this step only removes the free, but not the bound form of the negatively charged molecules. The bound form of these charged species is removed by passing the supernatant through a strong, high capacity anion exchange resin-containing column. An example of such a resin is the Macro-Prep™ High Q resin (Bio-Rad, Melville, N.Y.). Other strong, high capacity anion exchange columns are also suitable. The negatively charged molecules bind to the column, while the enzyme passes through the column. It is also found that some unrelated, undesirable proteins also bind to the column.

Next, the eluate from the anion exchange column is directly loaded to a cation exchange resin-

- 79 -

containing column. Examples of such resins are the S-Sepharose™ (Pharmacia, Piscataway, N.J.) and the Macro-Prep™ High S (Bio-Rad). Each of these two resin-containing columns has SO_3^- ligands bound thereto in order to facilitate the exchange of cations. Other cation exchange columns are also suitable. The enzyme binds to the column and is then eluted with a solvent capable of releasing the enzyme from the column.

Any salt which increases the conductivity of the solution is suitable for elution. Examples of such salts include sodium salts, as well as potassium salts and ammonium salts. An aqueous sodium chloride solution of appropriate concentration is suitable. A gradient, such as 0 to 250 mM sodium chloride is acceptable, as is a step elution using 200 mM sodium chloride.

A sharp peak is seen in the sodium chloride gradient elution (Figure 3). The improvement in enzyme yield over the prior method is striking. The recombinant chondroitinase I enzyme is recovered at a purity of 99% at a yield of 80-90%.

The purity of the protein is measured by scanning the bands in SDS-PAGE gels. A 4-20% gradient of acrylamide is used in the development of the gels. The band(s) in each lane of the gel is scanned using the procedure described above.

These improvements are related directly to the increase in binding of the enzyme to the cation exchange column which results from first using the anion exchange column. In comparative experiments, when only the cation exchange column is used, only 1% of the enzyme binds to the column. However, when the anion exchange column is used first, over 95% of the enzyme binds to the column.

35

- 80 -

Example 10

Second Method For The Isolation And
Purification Of Recombinant Chondroitinase I
According To This Invention

5

In the second embodiment of this aspect of the invention, two additional steps are inserted in the method before the diafiltration step of the first embodiment. The supernatant is treated with an acidic solution, such as 1 M acetic acid, bringing the supernatant to a final pH of 4.5, to precipitate out the desired enzyme. The pellet is obtained by centrifugation at 5,000 x g for 20 minutes. The pellet is then dissolved in an alkali solution, such as 20-30 mM NaOH, bringing it to a final pH of 9.8. The solution is then subjected to the diafiltration and subsequent steps of the first embodiment of this invention.

In comparative experiments with the second embodiment of this invention, when only the cation exchange column is used, only 5% of the enzyme binds to the column. However, when the anion exchange column is used first, essentially 100% of the enzyme binds to the column. The second embodiment provides comparable enzyme purity and yield to the first embodiment of the invention.

Acid precipitation removes proteins that remain soluble; however, these proteins are removed anyway by the cation and anion exchange steps that follow (although smaller columns may be used). An advantage of the acid precipitation step is that the sample volume is decreased to about 20% of the original volume after dissolution, and hence can be handled more easily on a large scale. However, the additional acid precipitation and alkali dissolution

- 81 -

steps of the second embodiment mean that the second embodiment is more time consuming than the first embodiment. On a manufacturing scale, the marginal improvements in purity and yield provided by the second embodiment may be outweighed by the simpler procedure of the first embodiment, which still provides highly pure enzyme at high yields.

The high purity of the recombinant enzyme obtained by the two embodiments of this invention is depicted in Figure 4. A single sharp band is seen in the SDS-PAGE gel photograph: Lane 1 is the enzyme using the method of the first embodiment; Lane 2 is the enzyme using the method of the second embodiment; Lane 3 represents the supernatant from the host cell prior to purification -- many other proteins are present; and Lane 4 represents molecular weight standards.

Example 11

20 Site-Specific Mutagenesis Of A Fragment Encoding The N-Terminal Region Of Chondroitinase II

The approach taken in the case of the chondroitinase II gene is to modify the naturally-occurring ATG initiation codon to embed it within an NdeI site. This results in a construction in which the signal peptide is retained, such that the expressed gene is processed and secreted to yield the mature native enzyme structure that has a leucine residue at the N-terminus. The mutagenized bases are upstream of the coding region.

The method used for this site-specific alteration is that described above for the expression of the chondroitinase I gene and is based on the work of Kunkel (15) using the Muta-Gene™ In Vitro

- 82 -

Mutagenesis Kit Version 2 (Bio-Rad, Melville, N.Y.). In this procedure, the target DNA to be mutagenized is first cloned into a suitable M13-derived vector to generate single-stranded DNA. This recombinant phage is replicated in the *E. coli* host strain CJ236 (Bio-Rad), a male strain that carries the dut and ung alleles. The combination of these two mutations, dut (duTPase) and ung (uracil-N-glycosylase), results in the incorporation of some uracil, rather than thymine, residues into the DNA synthesized in this organism. Single-stranded template is then isolated after propagation on CJ236 and the appropriate mutagenic, synthetic oligonucleotide (SEQ ID NO:41) is annealed to this DNA.

This oligonucleotide serves as a primer for T7 DNA polymerase which copies the entire recombinant molecule. T4 DNA ligase is then used to seal the nick between the first residue of the mutagenic oligonucleotide and the last residue added in vitro. The newly synthesized DNA (containing the desired base changes) therefore does not contain uracil while the template DNA (with the native sequences) does. Transformation of a non-mutant (with respect to the ung and dut alleles) male *E. coli* strain yields phage progeny that are primarily derived from the mutagenized strand synthesized in vitro as a result of the inactivation of the uracil-containing template strand.

In this specific case, the fragment to be cloned for the mutagenesis is a MunI-EcoRI fragment that spans the region between nucleotides 2943 to 3980 (SEQ ID NOS:1 and 39). The DNA digested to obtain this fragment is designated LP²783. This plasmid is constructed in the same way as LP²786 (described in Example 4), except that a HindIII linker is inserted

- 83 -

into the EcoRV deletion of LP²776 rather than the EcoRI linker. This MunI-EcoRI fragment is ligated into the unique EcoRI site of LP²941, an M13mp19 derivative in which the normal polylinker is replaced with that found in the plasmid vector pNEB193 (New England Biolabs, Beverly MA). The four base overhang produced by MunI digestion can be ligated to an EcoRI site, but the resulting recombinant sequence cannot be digested by either enzyme. The EcoRI digested LP²941 is also dephosphorylated with calf intestinal alkaline phosphatase (Boehringer Mannheim, Indianapolis IN) prior to gel purification and use.

The ligated DNA mixture is used to infect the male E. coli strain MV1190 and the plaques obtained are picked to 0.5 ml. of SM buffer and the phage allowed to elute by diffusion. These are then used to infect 10 ml. cultures of MV1190 and grown overnight. The cultures are centrifuged and the pellets used for the isolation of the double-stranded replicative forms of the recombinant virus. The supernatants, which contain the corresponding phage particles, are stored under refrigeration until needed. The orientation of the cloned fragment is determined by digestion of the replicative form DNA and HindIII, because there is one site within the polylinker and a second, asymmetrically placed site (SEQ ID NOS:1 and 39, nucleotides 3326-3331) within the above MunI-EcoRI fragment.

Once the desired orientation is identified, the corresponding phage-containing supernatant is serially diluted, used to infect the E. coli strain CJ236, and then plated to obtain single plaques which are picked and eluted as above. One of these is then used to infect CJ236 and another 10 ml culture grown and the single-stranded DNA is isolated from the

- 84 -

phage-containing supernatant using Qiaex™ columns and materials and methods recommended by the manufacturer (Qiagen, Chatsworth, CA) and finally resuspended in a volume of 0.01 ml. The recombinant phage are grown on
5 CJ236 (dut⁻ ung⁻) for two rounds in order to maximize the accumulation of uracil residues in the template and strand prior to the actual site-specific mutagenesis.

The mutagenic oligonucleotide used is
10 obtained from Bio-Synthesis (Denton, TX) and has the following sequence:

5'-ATT-TGC-AGG-AAA-TCT-GCA-TAT-GCT-AAT-AAA-AAA-CCC-3'
(SEQ ID NO:41)

15 This sequence differs from the corresponding region of SEQ ID NOS:1 and 39 in that an AT sequence (base pairs 3235 and 3236) is replaced by a CA sequence which creates the desired NdeI sequence
20 (CATATG) at the start of the presumed leader sequence for the chondroitinase II gene. One optical density unit of this oligonucleotide is dissolved in 0.46 ml. of TE 7.4 (0.01M TrisHCl, pH 7.8-0.001M EDTA, pH 8.0), yielding an oligonucleotide concentration of
25 approximately 6 pmol/μl. Three hundred picomoles of this oligonucleotide are phosphorylated in a 0.1 ml reaction containing 0.05 M TrisHCl, pH 7.8, 0.01 M MgCl₂, 0.02M dithiothreitol, 0.001 M ATP, 25 μg/ml bovine serum albumin and 100 units of T4
30 polynucleotide kinase (New England Biolabs) at 37°C for 30 minutes, followed by incubation at 75° for 20 minutes to inactivate the enzyme. The phosphorylated oligonucleotide is then stored frozen at -20° at a concentration of approximately 3 pmoles/μl.

35 For the site-specific mutagenesis, 1 μl (3

- 85 -

pmole) of the mutagenic oligonucleotide is mixed with 6 μ l of the single-stranded DNA prepared above in a 10 μ l volume of 0.02 M TrisHCl, pH 7.4, 0.002 M MgCl₂, 0.05 M NaCl. The oligonucleotide is annealed to this template by first incubating the sample at 70°C for 5 minutes and then cooling this sample at 25°C over a 45 minute period in a DNA Thermal Cycler™ (Perkin-Elmer Cetus/Norwalk, CT). The sample is maintained at 25°C for another 5 minutes before being cooled to 20°C and finally transferred to an ice bath.

The annealed primer is then extended after the addition of 1 μ l of 10X synthesis buffer (Bio-Rad; containing 0.005 M of each of the dNTP's, 0.01 M ATP, 0.1 M TrisHCl, pH 7.4, 0.05 M MgCl₂, 0.02 M DTT). One μ l of T4 DNA ligase (3 units/ μ l Bio-Rad) and 1 μ l of T7 DNA polymerase (0.5 units/ μ l Bio-Rad). The in vitro DNA synthesis is carried out on ice for 5 minutes, at 11°C for ten minutes, and at 37°C for 30 minutes prior to transfer to ice.

This sample is used directly to transform the male E. coli host MV1190 (dut⁺ ung⁺) and the resulting plaques, containing the site-specifically mutagenized phage, are obtained, picked and eluted as described above. Aliquots of these phage stocks are used in infect 10 ml. cultures of MV1190 and allowed to grow overnight. The cultures are centrifuged and the replicative forms of the recombinant phage are isolated using Qiaex™ columns and methods recommended by the manufacturer (Qiagen, Chatsworth CA). The DNA isolated is resuspended in 0.1 ml of TE 7.4. Initial digestions of a portion of each of these DNA samples with NdeI reveals that at least four appeared to have acquired a new NdeI site, indicate that the site-specific mutagenesis is successful. Consequently, larger samples of these four clones (0.04 ml each) are

- 86 -

digested with NdeI and EcoRI and fractionated on a 1.4% agarose gel run in a Tris-acetate-EDTA buffer system.

The desired approximately 740 base pair fragment is observed in each case and this band is excised from each pattern. The four samples are then combined and the DNA extracted from the gel using a Qiaex™ resin and buffers according to the manufacturer's recommendations (Qiagen, Chatsworth CA) and resuspended in 0.05 ml. of TE, pH 7.4. This isolated, site-specifically mutagenized N-terminal coding region of the cloned P. vulgaris gene for the chondroitinase II gene is then subcloned into the plasmid pNEB193 (New England Biolabs, Beverly MA) between the (dephosphorylated) unique NdeI and EcoRI sites present in this plasmid. After transformation of the E. coli host strain 294, 10 ml cultures derived from the individual transformants are grown and the recombinant plasmid DNA isolated as above. The DNA sample from one of the positive clones is designated m#15-5712. This sample represents the modified N-terminal region that is to be joined to the C-terminal coding region for the chondroitinase II gene, which is described in Example 12.

Example 12

Isolation, Characterization And DNA Sequence Analysis Of A Fragment Encoding The C-terminal Region Of Chondroitinase II

The DNA sequence contained in SEQ ID NO: 39 indicates that chondroitinase II is encoded by a region that is downstream of that for chondroitinase I. This information is derived from a portion of a 10 kilobase NsiI fragment of P. vulgaris that is

- 87 -

subcloned originally from a cosmid clone designated LP²751. The combination of the DNA sequencing and the restriction map in Figure 1 reveals that the chondroitinase II coding region initiates to the "left" of the EcoRI site that lies within the P. vulgaris derived DNA and proceeds toward the NsiI site at the "right" end of the fragment depicted in Figure 1. Therefore, this restriction map should be expanded to the "right" to find a suitable fragment that will include the C-terminal coding region for the chondroitinase II gene.

Digestion of LP²751 reveals three EcoRI fragments of approximately 20 kb, 13 kb, and 10 kb, and indicates that there are three EcoRI sites within LP²751. Because there are two EcoRI sites that bracket the cloning site, the conclusion is that there is one EcoRI site within the cloned P. vulgaris DNA in this clone. Furthermore, since the approximately 13 kb fragment corresponds to the size of the cosmid vector per se, this unique EcoRI site lies between the approximately 20 kb and the approximately 10 kb fragments noted above. Since it is known that the entire coding region for chondroitinase I, as well as the N-terminal coding region for chondroitinase II, are both contained within the approximately 10 kb NsiI fragment, restriction digestions that compare the patterns obtained among the cloned 10 kb NsiI (present in the recombinant plasmid designated LP²770) and gel-purified samples of the above approximately 20 kb EcoRI and approximately 10 kb EcoRI fragments indicate which of these EcoRI fragments contain the chondroitinase I coding sequence and, therefore by deduction, which will carry the C-terminal coding region for chondroitinase II. Consequently, digestions are carried using the restriction enzymes

- 88 -

AflIII, ClaI, EcoRV, and HindIII each of which has been noted by Applicants to yield eight to ten fragments upon digestion of the original cosmid clone designated LP²751.

5 The recombinant molecule carrying the subcloned approximately 10 kb NsiI fragment (LP²770) and the individually gel-purified approximately 20 kb EcoRI and approximately 10 kb EcoRI fragments are digested with each of these enzymes to yield patterns
10 of fragments that are compared. These digestions reveal that the approximately 20 kb EcoRI and the LP²770 patterns have a number of fragments in common. This indicates that the chondroitinase I gene and the N-terminal coding region of the chondroitinase II gene
15 are contained within the larger EcoRI fragment and, therefore, the C-terminal coding region for the chondroitinase II gene is on the approximately 10 kb EcoRI fragment.

 The approximately 10 kb EcoRI fragment is
20 cloned into the unique EcoRI site of the derivative of pNEB193 (New England Biolabs, Beverly MA) that is designated lacpoΔ pNEB193. This vector carries two deletions relative to the parental molecule pNEB193. The first removes the sequences between the unique
25 NdeI and EcoRI sites, retaining the EcoRI site but removing the NdeI site (and one of the two PvuII sites). The second deletion removes the region between the HindIII site at the other end of the polylinker and the (now unique) PvuII site,
30 maintaining the HindIII site, while removing the PvuII site. The recombinant DNA molecule carrying the subcloned approximately 10 kb EcoRI fragment in the vector lacpoΔ pNEB193 is designated LP²1263. The orientation of the 112 kD C-terminal coding region
35 within LP²1263 is determined by restriction enzyme

- 89 -

mapping. The results indicate that this region is positioned so as to proceed from the EcoRI site (defined as the "left" end) toward the HindIII site at the other end of the polylinker. Similarly, unique restriction sites for SmaI, XhoI, NocI and NdeI are found approximately 2.6, 4.6, 5.8 and 8.5 kb from the "left" end of the approximately 10 kb EcoRI fragment. Digestion of LP²1263 with SmaI, therefore, deletes a downstream region of approximately 7.4 kb from the site within the cloned P. vulgaris DNA to the second site within the polylinker region, leaving approximately 2.6 kb which should be enough to encode the missing region of the chondroitinase II gene. This construction also "places" a BamHI site (present in the polylinker) downstream of the coding region for the chondroitinase II gene. This recombinant DNA molecule which carries the chondroitinase II gene from the EcoRI site to (and presumably just beyond) the termination codon for this gene has been designated m#25-5712.

DNA sequence analysis is initiated on the approximately 10 kb EcoRI fragment derived from LP²1263 and is completed after the assembly of the intact gene for chondroitinase II. The materials and methods for the DNA sequencing of this fragment are essentially the same as those used for the approximately 4 kb fragment containing the gene for chondroitinase I. Random fragments are derived from this approximately 10 kb EcoRI fragment by self-ligating the DNA and then fragmenting the polymerized DNA by sonication as well as by partial digestion with the restriction enzymes Sau3A or MseI. These pieces are then eventually cloned into M13 derived vectors and the single-stranded recombinant molecules sequenced using the standard protocols described above.

- 90 -

Finally, with the two set of sequence data available, an approximately 300 base-pair BclI fragment is identified that is predicted to contain the EcoRI site that is the junction between the two P. vulgaris fragments of approximately 20 kb and approximately 10 kb obtained by digestion with EcoRI. This small fragment is sequenced in both directions to verify the nucleotide sequence through this junction point used in the constructions described below.

Example 13

Assembly Of The Entire Site-Specifically Mutagenized Gene For Chondroitinase II

During the DNA sequencing, the molecule designated m#25-5712 is digested with EcoRI and BamHI. This releases a DNA fragment of approximately 2.6 kb. Similarly, the construction designated m#15-5712 is digested with EcoRI and BamHI and then dephosphorylated prior to purification by gel electrophoresis. The latter molecule therefore carries the N-terminal coding region of the chondroitinase II gene from the ATG initiation codon (now present as part of an NdeI site from the site-specific mutagenesis) to the EcoRI site.

These two fragments are ligated and then the mixture used to transform the E. coli strain 294. Plasmid DNA is isolated from the transformants and positive clones identified. Restriction digestion with NdeI and BamHI releases the desired fragment encoding the chondroitinase II gene (SEQ ID NO:39, nucleotides 3235-6518, followed by 14 nucleotides derived from the polylinker, which includes a BamHI site). This fragment is then ligated to the expression vector pET9A (Novagen, Madison, WI)

- 91 -

described in the expression of the chondroitinase I gene.

The coding region of the chondroitinase II gene includes nucleotides 3238-6276 of the SEQ ID NO: 39, which encodes 1013 amino acids (SEQ ID NO:40). Of this region, nucleotides 3238-3306 encode the 23 amino acid signal peptide (SEQ ID NO:40, amino acids 1-23), while nucleotides 3307-6276 encode the mature 990 amino acid chondroitinase II protein (SEQ ID NO:40, amino acids 24-1013).

Restriction analysis with four enzymes of the region spanning both chondroitinase genes and flanking sequences thereof reveals the following restriction sites:

<u>Enzyme</u>	<u>Nucleotide</u>	<u>Enzyme</u>	<u>Nucleotide</u>
<u>EcoRI</u>	2	<u>MunI</u>	4510
<u>HindIII</u>	2046	<u>HindIII</u>	4530
<u>MunI</u>	2904	<u>MunI</u>	5176
<u>MunI</u>	2943	<u>HindIII</u>	5427
<u>HindIII</u>	3326	<u>SmaI</u>	6515
<u>EcoRI</u>	3974		

In addition, restriction analysis with Sau3AI reveals a multiplicity of sites, including those at SEQ ID NO:39, nucleotides 212, 602, 890, 1042, 1181, 1241, 1442, 1505, 1746, 2330, 2363, 2701, 2705, 2920, 3697, 3708, 3745, 3868, 4087, 4800, 4872, 5565, 5635, 5860, 6058 and 6467.

One of the recombinant molecules (the chondroitinase II gene inserted into pET9A) obtained in this experiment is grown in large scale (0.5 liter) and the expression system containing the chondroitinase II gene isolated and designated LP²1359. An aliquot of this DNA is used to transform the expression host BL21(DE3)/pLysS described in the

- 92 -

expression of the chondroitinase I gene. The resulting strain is designated TD112 and is used for large-scale fermentation and isolation of the chondroitinase II enzyme.

5 A fermentation at a 10 liter scale carried out with this E. coli strain containing the plasmid expressing the chondroitinase II protein, provides a maximum chondroitinase II titer of approximately 0.3 mg/ml, which is approximately 25 times that of the
10 approximately 0.012 mg/ml obtained from the native P. vulgaris fermentation process for chondroitinase II.

Example 14

15 First Method For The Isolation And Purification Of Recombinant Chondroitinase II
According To This Invention

The initial part of this method is the same as that used for the recombinant chondroitinase I
20 enzyme. As a first step, the host cells which express the recombinant chondroitinase II enzyme are homogenized to lyse the cells. This releases the enzyme into the supernatant.

In one embodiment of this invention, the
25 supernatant is first subjected to diafiltration to remove salts and other small molecules. An example of a suitable filter is a spiral wound 30 kD filter made by Amicon (Beverly, MA). However, this step only removes the free, but not the bound form of the
30 negatively charged molecules. The bound form of these charged species is removed by passing the supernatant (see the SDS-PAGE gel depicted in Figure 5, lane 1) through a strong, high capacity anion exchange resin-containing column. An example of such a resin is the
35 Macro-Prep™ High Q resin (Bio-Rad, Melville, N.Y.).

- 93 -

Other strong, high capacity anion exchange columns are also suitable. The negatively charged molecules bind to the column, while the enzyme passes through the column with approximately 90% recovery of the enzyme. It is also found that some unrelated, undesirable proteins also bind to the column.

Next, the eluate from the anion exchange column (Figure 5, lane 2) is directly loaded to a cation exchange resin-containing column. Examples of such resins are the S-Sepharose™ (Pharmacia, Piscataway, N.J.) and the Macro-Prep™ High S (Bio-Rad). Each of these two resin-containing columns has SO_3^- ligands bound thereto in order to facilitate the exchange of cations. Other cation exchange columns are also suitable. The enzyme binds to the column, while a significant portion of contaminating proteins elute unbound.

At this point, the method diverges from that used for the chondroitinase I protein. Instead of eluting the protein with a non-specific salt solution capable of releasing the enzyme from the cation exchange column, a specific elution using a solution containing chondroitin sulfate is used. A 1% concentration of chondroitin sulfate is used; however, a gradient of this solvent is also acceptable. The specific chondroitin sulfate solution is preferred to the non-specific salt solution because the recombinant chondroitinase II protein is expressed at levels approximately several-fold lower than the recombinant chondroitinase I protein; therefore, a more powerful and selective solution is necessary in order to obtain a final chondroitinase II product of a purity equivalent to that obtained for the chondroitinase I protein.

The cation exchange column is next washed

- 94 -

with a phosphate buffer, pH 7.0, to elute unbound proteins, followed by washing with borate buffer, pH 8.5, to elute loosely bound contaminating proteins and to increase the pH of the resin to that required for the optimal elution of the chondroitinase II protein using the substrate, chondroitin sulfate.

Next, a 1% solution of chondroitin sulfate in water, adjusted to pH 9.0, is used to elute the chondroitinase II protein, as a sharp peak (recovery 65%) and at a high purity of approximately 95% (Figure 5, lane 3). However, the chondroitin sulfate has an affinity for the chondroitinase II protein which is stronger than its affinity for the resin of the column, and therefore the chondroitin sulfate co-elutes with the protein. This ensures that only protein which recognizes chondroitin sulfate is eluted, which is desirable, but also means that an additional process step is necessary to separate the chondroitin sulfate from the chondroitinase II protein.

In this separation step, the eluate is adjusted to pH 7.0 and is loaded as is onto an anion exchange resin-containing column, such as the Macro-Prep™ High Q resin. The column is washed with a 20 mM phosphate buffer, pH 6.8. The chondroitin sulfate binds to the column, while the chondroitinase II protein flows through in the unbound pool with greater than 95% recovery. At this point, the protein is pure, except for the presence of a single minor contaminant of approximately 37 kD (Figure 5, lanes 4 and 6). The contaminant may be a breakdown product of the chondroitinase II protein.

This contaminant is effectively removed by a crytallization step. The eluate from the anion exchange column is concentrated to 15 mg/ml protein

- 95 -

using an Amicon stirred cell with a 30 kD cutoff. The solution is maintained at 4°C for several days to crystallize out the pure chondroitinase II protein. The supernatant contains the 37 kD contaminant (Figure 5, lane 7). Centrifugation causes the crystals to form a pellet, while the supernatant with the 37 kD contaminant is removed by pipetting, and the crystals washed twice with water. After the first wash, some of the contaminant remains (Figure 5, lane 8), but after the second wash, only the chondroitinase II protein is visible (Figure 5, lane 9). The washed crystals are redissolved in water and exhibit a single protein band on SDS-PAGE, with a purity of greater than 99% (Figure 5, lane 10).

Example 15

Second Method For The Isolation And Purification Of Recombinant Chondroitinase II According To This Invention

In the second embodiment of this aspect of the invention, two additional steps are inserted in the method for purifying the chondroitinase II enzyme before the diafiltration step of the first embodiment. The supernatant is treated with an acidic solution, such as 1 M acetic acid, bringing the supernatant to a final pH of 4.5, to precipitate out the desired enzyme. The pellet is obtained by centrifugation at 5,000 x g for 20 minutes. The pellet is then dissolved in an alkali solution, such as 20-30 mM NaOH, bringing it to a final pH of 9.8. The solution is then subjected to the diafiltration and subsequent steps of the first embodiment of this aspect of the invention.

- 96 -

Bibliography

1. Yamagata, T., et al., J. Biol. Chem., 243, 1523-1535 (1968).
2. Kikuchi, H., et al., U.S. Patent Number 5,198,355.
3. Brown, M. D., U.S. Patent Number 4,696,816.
4. Hageman, G. S., U.S. Patent Number 5,292,509.
5. Malitschek, B., and Scharf, M., Biotechniques, 7, 177-178 (1991).
6. Sanger, F., et al., Proc. Natl. Acad. Sci. USA, 74, 5463-5467 (1977).
7. Innis, M. A., and Gelfand, D. H., "Optimization of PCRs", pages 3-12 in PCR Protocols, A Guide to Methods and Applications, Academic Press, New York, N.Y. (1990).
8. Southern, E., J. Mol. Biol., 98, 503-517 (1975).
9. Studier, F. W., et al., Methods in Enzymology, 185, 60-89 (1990).
10. Studier, F. W., and Moffatt, B. A., J. Mol. Biol., 189, 113-130 (1986); Moffatt, B. A., and Studier, F. W., Cell, 49, 221-227 (1987).
11. Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).
12. Clark, Nucleic Acids Research, 16, 9677 (1988).
13. Yanisch-Perron, C., et al., Gene, 3, 103-119 (1985).
14. Devereaux, et al., Nucleic Acids Research, 12, 387-395 (1984).
15. Kunkel, T. A., Proc. Natl. Acad. Sci.,

- 97 -

USA, 82, 488-492 (1985).

16. Studier, F. W., J. Mol. Biol., 219, 37-44 (1991).

17. Miller, J. H., Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1972).

- 98 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: American Cyanamid Company
- (ii) TITLE OF INVENTION: Cloning And Expression Of The Chondroitinase I and II Genes From P. Vulgaris
- (iii) NUMBER OF SEQUENCES: 41
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: American Cyanamid Company
 - (B) STREET: One Cyanamid Plaza
 - (C) CITY: Wayne
 - (D) STATE: New Jersey
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 07470-8426
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US94/
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Gordon, Alan M.
 - (B) REGISTRATION NUMBER: 30,637
 - (C) REFERENCE/DOCKET NUMBER: 31,726-00/PCT
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 201-831-3244
 - (B) TELEFAX: 201-831-3305

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3980 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 119..3181
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

SUBSTITUTE SHEET (RULE 26)

- 99 -

GGAATTCCAT CACTCAATCA TTAAATTTAG GCACAACGAT GGGCTATCAG CGTTATGACA	60
AATTTAATGA AGGACGCATT GGTTTCACTG TTAGCCAGCG TTTCTAAGGA GAAAAATA	118
ATG CCG ATA TTT CGT TTT ACT GCA CTT GCA ATG ACA TTG GGG CTA TTA Met Pro Ile Phe Arg Phe Thr Ala Leu Ala Met Thr Leu Gly Leu Leu 1 5 10 15	166
TCA GCG CCT TAT AAC GCG ATG GCA GCC ACC AGC AAT CCT GCA TTT GAT Ser Ala Pro Tyr Asn Ala Met Ala Ala Thr Ser Asn Pro Ala Phe Asp 20 25 30	214
CCT AAA AAT CTG ATG CAG TCA GAA ATT TAC CAT TTT GCA CAA AAT AAC Pro Lys Asn Leu Met Gln Ser Glu Ile Tyr His Phe Ala Gln Asn Asn 35 40 45	262
CCA TTA GCA GAC TTC TCA TCA GAT AAA AAC TCA ATA CTA ACG TTA TCT Pro Leu Ala Asp Phe Ser Ser Asp Lys Asn Ser Ile Leu Thr Leu Ser 50 55 60	310
GAT AAA CGT AGC ATT ATG GGA AAC CAA TCT CTT TTA TGG AAA TGG AAA Asp Lys Arg Ser Ile Met Gly Asn Gln Ser Leu Leu Trp Lys Trp Lys 65 70 75 80	358
GGT GGT AGT AGC TTT ACT TTA CAT AAA AAA CTG ATT GTC CCC ACC GAT Gly Gly Ser Ser Phe Thr Leu His Lys Lys Leu Ile Val Pro Thr Asp 85 90 95	406
AAA GAA GCA TCT AAA GCA TGG GGA CGC TCA TCT ACC CCC GTT TTC TCA Lys Glu Ala Ser Lys Ala Trp Gly Arg Ser Ser Thr Pro Val Phe Ser 100 105 110	454
TTT TGG CTT TAC AAT GAA AAA CCG ATT GAT GGT TAT CTT ACT ATC GAT Phe Trp Leu Tyr Asn Glu Lys Pro Ile Asp Gly Tyr Leu Thr Ile Asp 115 120 125	502
TTC GGA GAA AAA CTC ATT TCA ACC AGT GAG GCT CAG GCA GGC TTT AAA Phe Gly Glu Lys Leu Ile Ser Thr Ser Glu Ala Gln Ala Gly Phe Lys 130 135 140	550
GTA AAA TTA GAT TTC ACT GGC TGG CGT GCT GTG GGA GTC TCT TTA AAT Val Lys Leu Asp Phe Thr Gly Trp Arg Ala Val Gly Val Ser Leu Asn 145 150 155 160	598
AAC GAT CTT GAA AAT CGA GAG ATG ACC TTA AAT GCA ACC AAT ACC TCC Asn Asp Leu Glu Asn Arg Glu Met Thr Leu Asn Ala Thr Asn Thr Ser 165 170 175	646
TCT GAT GGT ACT CAA GAC AGC ATT GGG CGT TCT TTA GGT GCT AAA GTC Ser Asp Gly Thr Gln Asp Ser Ile Gly Arg Ser Leu Gly Ala Lys Val 180 185 190	694
GAT AGT ATT CGT TTT AAA GCG CCT TCT AAT GTG AGT CAG GGT GAA ATC Asp Ser Ile Arg Phe Lys Ala Pro Ser Asn Val Ser Gln Gly Glu Ile 195 200 205	742
TAT ATC GAC CGT ATT ATG TTT TCT GTC GAT GAT GCT CGC TAC CAA TGG Tyr Ile Asp Arg Ile Met Phe Ser Val Asp Asp Ala Arg Tyr Gln Trp 210 215 220	790
TCT GAT TAT CAA GTA AAA ACT CGC TTA TCA GAA CCT GAA ATT CAA TTT Ser Asp Tyr Gln Val Lys Thr Arg Leu Ser Glu Pro Glu Ile Gln Phe	838

SUBSTITUTE SHEET (RULE 26)

- 100 -

225					230					235					240	
CAC	AAC	GTA	AAG	CCA	CAA	CTA	CCT	GTA	ACA	CCT	GAA	AAT	TTA	GCG	GCC	886
His	Asn	Val	Lys	Pro	Gln	Leu	Pro	Val	Thr	Pro	Glu	Asn	Leu	Ala	Ala	
				245					250					255		
ATT	GAT	CTT	ATT	CGC	CAA	CGT	CTA	ATT	AAT	GAA	TTT	GTC	GGA	GGT	GAA	934
Ile	Asp	Leu	Ile	Arg	Gln	Arg	Leu	Ile	Asn	Glu	Phe	Val	Gly	Gly	Glu	
			260					265					270			
AAA	GAG	ACA	AAC	CTC	GCA	TTA	GAA	GAG	AAT	ATC	AGC	AAA	TTA	AAA	AGT	982
Lys	Glu	Thr	Asn	Leu	Ala	Leu	Glu	Glu	Asn	Ile	Ser	Lys	Leu	Lys	Ser	
		275					280					285				
GAT	TTC	GAT	GCT	CTT	AAT	ATT	CAC	ACT	TTA	GCA	AAT	GGT	GGA	ACG	CAA	1030
Asp	Phe	Asp	Ala	Leu	Asn	Ile	His	Thr	Leu	Ala	Asn	Gly	Gly	Thr	Gln	
	290					295					300					
GGC	AGA	CAT	CTG	ATC	ACT	GAT	AAA	CAA	ATC	ATT	ATT	TAT	CAA	CCA	GAG	1078
Gly	Arg	His	Leu	Ile	Thr	Asp	Lys	Gln	Ile	Ile	Ile	Tyr	Gln	Pro	Glu	
305				310					315						320	
AAT	CTT	AAC	TCC	CAA	GAT	AAA	CAA	CTA	TTT	GAT	AAT	TAT	GTT	ATT	TTA	1126
Asn	Leu	Asn	Ser	Gln	Asp	Lys	Gln	Leu	Phe	Asp	Asn	Tyr	Val	Ile	Leu	
				325					330					335		
GGT	AAT	TAC	ACG	ACA	TTA	ATG	TTT	AAT	ATT	AGC	CGT	GCT	TAT	GTG	CTG	1174
Gly	Asn	Tyr	Thr	Thr	Leu	Met	Phe	Asn	Ile	Ser	Arg	Ala	Tyr	Val	Leu	
			340				345						350			
GAA	AAA	GAT	CCC	ACA	CAA	AAG	GCG	CAA	CTA	AAG	CAG	ATG	TAC	TTA	TTA	1222
Glu	Lys	Asp	Pro	Thr	Gln	Lys	Ala	Gln	Leu	Lys	Gln	Met	Tyr	Leu	Leu	
		355					360					365				
ATG	ACA	AAG	CAT	TTA	TTA	GAT	CAA	GGC	TTT	GTT	AAA	GGG	AGT	GCT	TTA	1270
Met	Thr	Lys	His	Leu	Leu	Asp	Gln	Gly	Phe	Val	Lys	Gly	Ser	Ala	Leu	
	370					375					380					
GTG	ACA	ACC	CAT	CAC	TGG	GGA	TAC	AGT	TCT	CGT	TGG	TGG	TAT	ATT	TCC	1318
Val	Thr	Thr	His	His	Trp	Gly	Tyr	Ser	Ser	Arg	Trp	Trp	Tyr	Ile	Ser	
385					390					395					400	
ACG	TTA	TTA	ATG	TCT	GAT	GCA	CTA	AAA	GAA	GCG	AAC	CTA	CAA	ACT	CAA	1366
Thr	Leu	Leu	Met	Ser	Asp	Ala	Leu	Lys	Glu	Ala	Asn	Leu	Gln	Thr	Gln	
				405					410					415		
GTT	TAT	GAT	TCA	TTA	CTG	TGG	TAT	TCA	CGT	GAG	TTT	AAA	AGT	AGT	TTT	1414
Val	Tyr	Asp	Ser	Leu	Leu	Trp	Tyr	Ser	Arg	Glu	Phe	Lys	Ser	Ser	Phe	
			420					425					430			
GAT	ATG	AAA	GTA	AGT	GCT	GAT	AGC	TCT	GAT	CTA	GAT	TAT	TTC	AAT	ACC	1462
Asp	Met	Lys	Val	Ser	Ala	Asp	Ser	Ser	Asp	Leu	Asp	Tyr	Phe	Asn	Thr	
		435					440					445				
TTA	TCT	CGC	CAA	CAT	TTA	GCC	TTA	TTA	TTA	CTA	GAG	CCT	GAT	GAT	CAA	1510
Leu	Ser	Arg	Gln	His	Leu	Ala	Leu	Leu	Leu	Leu	Glu	Pro	Asp	Asp	Gln	
	450					455					460					
AAG	CGT	ATC	AAC	TTA	GTT	AAT	ACT	TTC	AGC	CAT	TAT	ATC	ACT	GGC	GCA	1558
Lys	Arg	Ile	Asn	Leu	Val	Asn	Thr	Phe	Ser	His	Tyr	Ile	Thr	Gly	Ala	
465					470					475					480	

SUBSTITUTE SHEET (RULE 26)

- 101 -

TTA	ACG	CAA	GTG	CCA	CCG	GGT	GGT	AAA	GAT	GGT	TTA	CGC	CCT	GAT	GGT	1606
Leu	Thr	Gln	Val	Pro	Pro	Gly	Gly	Lys	Asp	Gly	Leu	Arg	Pro	Asp	Gly	
				485					490					495		
ACA	GCA	TGG	CGA	CAT	GAA	GGC	AAC	TAT	CCG	GGC	TAC	TCT	TTC	CCA	GCC	1654
Thr	Ala	Trp	Arg	His	Glu	Gly	Asn	Tyr	Pro	Gly	Tyr	Ser	Phe	Pro	Ala	
			500					505					510			
TTT	AAA	AAT	GCC	TCT	CAG	CTT	ATT	TAT	TTA	TTA	CGC	GAT	ACA	CCA	TTT	1702
Phe	Lys	Asn	Ala	Ser	Gln	Leu	Ile	Tyr	Leu	Leu	Arg	Asp	Thr	Pro	Phe	
		515					520					525				
TCA	GTG	GGT	GAA	AGT	GGT	TGG	AAT	AAC	CTG	AAA	AAA	GCG	ATG	GTT	TCA	1750
Ser	Val	Gly	Glu	Ser	Gly	Trp	Asn	Asn	Leu	Lys	Lys	Ala	Met	Val	Ser	
	530					535					540					
GCG	TGG	ATC	TAC	AGT	AAT	CCA	GAA	GTT	GGA	TTA	CCG	CTT	GCA	GGA	AGA	1798
Ala	Trp	Ile	Tyr	Ser	Asn	Pro	Glu	Val	Gly	Leu	Pro	Leu	Ala	Gly	Arg	
545					550					555					560	
CAC	CCT	TTT	AAC	TCA	CCT	TCG	TTA	AAA	TCA	GTC	GCT	CAA	GGC	TAT	TAC	1846
His	Pro	Phe	Asn	Ser	Pro	Ser	Leu	Lys	Ser	Val	Ala	Gln	Gly	Tyr	Tyr	
				565					570					575		
TGG	CTT	GCC	ATG	TCT	GCA	AAA	TCA	TCG	CCT	GAT	AAA	ACA	CTT	GCA	TCT	1894
Trp	Leu	Ala	Met	Ser	Ala	Lys	Ser	Ser	Pro	Asp	Lys	Thr	Leu	Ala	Ser	
			580					585					590			
ATT	TAT	CTT	GCG	ATT	AGT	GAT	AAA	ACA	CAA	AAT	GAA	TCA	ACT	GCT	ATT	1942
Ile	Tyr	Leu	Ala	Ile	Ser	Asp	Lys	Thr	Gln	Asn	Glu	Ser	Thr	Ala	Ile	
		595					600					605				
TTT	GGA	GAA	ACT	ATT	ACA	CCA	GCG	TCT	TTA	CCT	CAA	GGT	TTC	TAT	GCC	1990
Phe	Gly	Glu	Thr	Ile	Thr	Pro	Ala	Ser	Leu	Pro	Gln	Gly	Phe	Tyr	Ala	
	610					615					620					
TTT	AAT	GGC	GGT	GCT	TTT	GGT	ATT	CAT	CGT	TGG	CAA	GAT	AAA	ATG	GTG	2038
Phe	Asn	Gly	Gly	Ala	Phe	Gly	Ile	His	Arg	Trp	Gln	Asp	Lys	Met	Val	
625					630					635					640	
ACA	CTG	AAA	GCT	TAT	AAC	ACC	AAT	GTT	TGG	TCA	TCT	GAA	ATT	TAT	AAC	2086
Thr	Leu	Lys	Ala	Tyr	Asn	Thr	Asn	Val	Trp	Ser	Ser	Glu	Ile	Tyr	Asn	
				645					650					655		
AAA	GAT	AAC	CGT	TAT	GGC	CGT	TAC	CAA	AGT	CAT	GGT	GTC	GCT	CAA	ATA	2134
Lys	Asp	Asn	Arg	Tyr	Gly	Arg	Tyr	Gln	Ser	His	Gly	Val	Ala	Gln	Ile	
			660					665					670			
GTG	AGT	AAT	GGC	TCG	CAG	CTT	TCA	CAG	GGC	TAT	CAG	CAA	GAA	GGT	TGG	2182
Val	Ser	Asn	Gly	Ser	Gln	Leu	Ser	Gln	Gly	Tyr	Gln	Gln	Glu	Gly	Trp	
		675					680					685				
GAT	TGG	AAT	AGA	ATG	CAA	GGG	GCA	ACC	ACT	ATT	CAC	CTT	CCT	CTT	AAA	2230
Asp	Trp	Asn	Arg	Met	Gln	Gly	Ala	Thr	Thr	Ile	His	Leu	Pro	Leu	Lys	
	690					695					700					
GAC	TTA	GAC	AGT	CCT	AAA	CCT	CAT	ACC	TTA	ATG	CAA	CGT	GGA	GAG	CGT	2278
Asp	Leu	Asp	Ser	Pro	Lys	Pro	His	Thr	Leu	Met	Gln	Arg	Gly	Glu	Arg	
705					710					715					720	
GGA	TTT	AGC	GGA	ACA	TCA	TCC	CTT	GAA	GGT	CAA	TAT	GGC	ATG	ATG	GCA	2326
Gly	Phe	Ser	Gly	Thr	Ser	Ser	Leu	Glu	Gly	Gln	Tyr	Gly	Met	Met	Ala	

SUBSTITUTE SHEET (RULE 26)

- 102 -

				725				730				735				
TTC	GAT	CTT	ATT	TAT	CCC	GCC	AAT	CTT	GAG	CGT	TTT	GAT	CCT	AAT	TTC	2374
Phe	Asp	Leu	Ile	Tyr	Pro	Ala	Asn	Leu	Glu	Arg	Phe	Asp	Pro	Asn	Phe	
740				745				750								
ACT	GCG	AAA	AAG	AGT	GTA	TTA	GCC	GCT	GAT	AAT	CAC	TTA	ATT	TTT	ATT	2422
Thr	Ala	Lys	Lys	Ser	Val	Leu	Ala	Ala	Asp	Asn	His	Leu	Ile	Phe	Ile	
755				760				765								
GGT	AGC	AAT	ATA	AAT	AGT	AGT	GAT	AAA	AAT	AAA	AAT	GTT	GAA	ACG	ACC	2470
Gly	Ser	Asn	Ile	Asn	Ser	Ser	Asp	Lys	Asn	Lys	Asn	Val	Glu	Thr	Thr	
770				775				780								
TTA	TTC	CAA	CAT	GCC	ATT	ACT	CCA	ACA	TTA	AAT	ACC	CTT	TGG	ATT	AAT	2518
Leu	Phe	Gln	His	Ala	Ile	Thr	Pro	Thr	Leu	Asn	Thr	Leu	Trp	Ile	Asn	
785				790				795				800				
GGA	CAA	AAG	ATA	GAA	AAC	ATG	CCT	TAT	CAA	ACA	ACA	CTT	CAA	CAA	GGT	2566
Gly	Gln	Lys	Ile	Glu	Asn	Met	Pro	Tyr	Gln	Thr	Thr	Leu	Gln	Gln	Gly	
805				810				815								
GAT	TGG	TTA	ATT	GAT	AGC	AAT	GGC	AAT	GGT	TAC	TTA	ATT	ACT	CAA	GCA	2614
Asp	Trp	Leu	Ile	Asp	Ser	Asn	Gly	Asn	Gly	Tyr	Leu	Ile	Thr	Gln	Ala	
820				825				830								
GAA	AAA	GTA	AAT	GTA	AGT	CGC	CAA	CAT	CAG	GTT	TCA	GCG	GAA	AAT	AAA	2662
Glu	Lys	Val	Asn	Val	Ser	Arg	Gln	His	Gln	Val	Ser	Ala	Glu	Asn	Lys	
835				840				845								
AAT	CGC	CAA	CCG	ACA	GAA	GGA	AAC	TTT	AGC	TCG	GCA	TGG	ATC	GAT	CAC	2710
Asn	Arg	Gln	Pro	Thr	Glu	Gly	Asn	Phe	Ser	Ser	Ala	Trp	Ile	Asp	His	
850				855				860								
AGC	ACT	CGC	CCC	AAA	GAT	GCC	AGT	TAT	GAG	TAT	ATG	GTC	TTT	TTA	GAT	2758
Ser	Thr	Arg	Pro	Lys	Asp	Ala	Ser	Tyr	Glu	Tyr	Met	Val	Phe	Leu	Asp	
865				870				875				880				
GCG	ACA	CCT	GAA	AAA	ATG	GGA	GAG	ATG	GCA	CAA	AAA	TTC	CGT	GAA	AAT	2806
Ala	Thr	Pro	Glu	Lys	Met	Gly	Glu	Met	Ala	Gln	Lys	Phe	Arg	Glu	Asn	
885				890				895								
AAT	GGG	TTA	TAT	CAG	GTT	CTT	CGT	AAG	GAT	AAA	GAC	GTT	CAT	ATT	ATT	2854
Asn	Gly	Leu	Tyr	Gln	Val	Leu	Arg	Lys	Asp	Lys	Asp	Val	His	Ile	Ile	
900				905				910								
CTC	GAT	AAA	CTC	AGC	AAT	GTA	ACG	GGA	TAT	GCC	TTT	TAT	CAG	CCA	GCA	2902
Leu	Asp	Lys	Leu	Ser	Asn	Val	Thr	Gly	Tyr	Ala	Phe	Tyr	Gln	Pro	Ala	
915				920				925								
TCA	ATT	GAA	GAC	AAA	TGG	ATC	AAA	AAG	GTT	AAT	AAA	CCT	GCA	ATT	GTG	2950
Ser	Ile	Glu	Asp	Lys	Trp	Ile	Lys	Lys	Val	Asn	Lys	Pro	Ala	Ile	Val	
930				935				940								
ATG	ACT	CAT	CGA	CAA	AAA	GAC	ACT	CTT	ATT	GTC	AGT	GCA	GTT	ACA	CCT	2998
Met	Thr	His	Arg	Gln	Lys	Asp	Thr	Leu	Ile	Val	Ser	Ala	Val	Thr	Pro	
945				950				955				960				
GAT	TTA	AAT	ATG	ACT	CGC	CAA	AAA	GCA	GCA	ACT	CCT					

SUBSTITUTE SHEET (RULE 26)

- 103 -

GTC ACG ATT AAT GGC AAA TGG CAA TCT GCT GAT AAA AAT AGT GAA GTG	3094
Val Thr Ile Asn Gly Lys Trp Gln Ser Ala Asp Lys Asn Ser Glu Val	
980 985 990	
AAA TAT CAG GTT TCT GGT GAT AAC ACT GAA CTG ACG TTT ACG AGT TAC	3142
Lys Tyr Gln Val Ser Gly Asp Asn Thr Glu Leu Thr Phe Thr Ser Tyr	
995 1000 1005	
TTT GGT ATT CCA CAA GAA ATC AAA CTC TCG CCA CTC CCT TGATTTAATC	3191
Phe Gly Ile Pro Gln Glu Ile Lys Leu Ser Pro Leu Pro	
1010 1015 1020	
AAAAGAACGC TCTTGC GTTC CTTTTTTTATT TGCAGGAAAT CTGATTATGC TAATAAAAAA	3251
CCCTTTAGCC CACGCGGTTA CATTAAAGCCT CTGTTTATCA TTACCCGCAC AAGCATTACC	3311
CACCTGTGCT CATGAAGCTT TCGGCGATAT TTATCTTTTT GAAGGTGAAT TACCCAATAC	3371
CCTTACCACT TCAAATAATA ATCAATTATC GCTAAGCAAA CAGCATGCTA AAGATGGTGA	3431
ACAATCACTC AAATGGCAAT ATCAACCACA AGCAACATTA AACTAAATA ATATTGTTAA	3491
TTACCAAGAT GATAAAAAATA CAGCCACACC ACTCACTTTT ATGATGTGGA TTTATAATGA	3551
AAAACCTCAA TCTTCCCAT TAACGTTAGC ATTTAAACAA AATAATAAAA TTGCACTAAG	3611
TTTTAATGCT GAACTTAATT TTACGGGGTG GCGAGGTATT GCTGTTCCTT TTCGTGATAT	3671
GCAAGGCTCT GCGACAGGTC AACTTGATCA ATTAGTGATC ACCGCTCCAA ACCAAGCCGG	3731
AACACTCTTT TTTGATCAAA TCATCATGAG TGTACCGTTA GACAATCGTT GGGCAGTACC	3791
TGACTATCAA ACACCTTACG TAAATAACGC AGTAAACACG ATGGTTAGTA AAAACTGGAG	3851
TGCATTATTG ATGTACGATC AGATGTTTCA AGCCCATTAC CCTACTTTAA ACTTCGATAC	3911
TGAATTTTCG GATGACCAA CAGAAATGGC TTCGATTTAT CAGCGCTTTG AATATTATCA	3971
AGGAATTCC	3980

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1021 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Ile	Phe	Arg	Phe	Thr	Ala	Leu	Ala	Met	Thr	Leu	Gly	Leu	Leu
1				5					10					15	
Ser	Ala	Pro	Tyr	Asn	Ala	Met	Ala	Ala	Thr	Ser	Asn	Pro	Ala	Phe	Asp
		20						25					30		
Pro	Lys	Asn	Leu	Met	Gln	Ser	Glu	Ile	Tyr	His	Phe	Ala	Gln	Asn	Asn
		35					40					45			
Pro	Leu	Ala	Asp	Phe	Ser	Ser	Asp	Lys	Asn	Ser	Ile	Leu	Thr	Leu	Ser

SUBSTITUTE SHEET (RULE 26)

- 104 -

50					55					60					
Asp 65	Lys	Arg	Ser	Ile	Met 70	Gly	Asn	Gln	Ser	Leu 75	Leu	Trp	Lys	Trp	Lys 80
Gly	Gly	Ser	Ser	Phe 85	Thr	Leu	His	Lys	Lys 90	Leu	Ile	Val	Pro	Thr	Asp 95
Lys	Glu	Ala	Ser 100	Lys	Ala	Trp	Gly	Arg 105	Ser	Ser	Thr	Pro	Val 110	Phe	Ser
Phe	Trp	Leu 115	Tyr	Asn	Glu	Lys	Pro 120	Ile	Asp	Gly	Tyr	Leu 125	Thr	Ile	Asp
Phe 130	Gly	Glu	Lys	Leu	Ile	Ser 135	Thr	Ser	Glu	Ala	Gln 140	Ala	Gly	Phe	Lys
Val 145	Lys	Leu	Asp	Phe	Thr 150	Gly	Trp	Arg	Ala	Val 155	Gly	Val	Ser	Leu	Asn 160
Asn	Asp	Leu	Glu	Asn 165	Arg	Glu	Met	Thr	Leu	Asn 170	Ala	Thr	Asn	Thr	Ser 175
Ser	Asp	Gly	Thr 180	Gln	Asp	Ser	Ile	Gly 185	Arg	Ser	Leu	Gly	Ala 190	Lys	Val
Asp	Ser	Ile 195	Arg	Phe	Lys	Ala	Pro 200	Ser	Asn	Val	Ser	Gln 205	Gly	Glu	Ile
Tyr 210	Ile	Asp	Arg	Ile	Met	Phe 215	Ser	Val	Asp	Asp	Ala 220	Arg	Tyr	Gln	Trp
Ser 225	Asp	Tyr	Gln	Val	Lys 230	Thr	Arg	Leu	Ser	Glu 235	Pro	Glu	Ile	Gln	Phe 240
His	Asn	Val	Lys	Pro 245	Gln	Leu	Pro	Val	Thr 250	Pro	Glu	Asn	Leu	Ala 255	Ala
Ile	Asp	Leu	Ile 260	Arg	Gln	Arg	Leu	Ile 265	Asn	Glu	Phe	Val	Gly 270	Gly	Glu
Lys	Glu	Thr 275	Asn	Leu	Ala	Leu	Glu 280	Glu	Asn	Ile	Ser	Lys 285	Leu	Lys	Ser
Asp	Phe 290	Asp	Ala	Leu	Asn	Ile 295	His	Thr	Leu	Ala	Asn 300	Gly	Gly	Thr	Gln
Gly 305	Arg	His	Leu	Ile	Thr 310	Asp	Lys	Gln	Ile	Ile 315	Ile	Tyr	Gln	Pro	Glu 320
Asn	Leu	Asn	Ser	Gln 325	Asp	Lys	Gln	Leu	Phe 330	Asp	Asn	Tyr	Val	Ile 335	Leu
Gly	Asn	Tyr	Thr 340	Thr	Leu	Met	Phe	Asn 345	Ile	Ser	Arg	Ala	Tyr	Val 350	Leu
Glu	Lys	Asp 355	Pro	Thr	Gln	Lys	Ala 360	Gln	Leu	Lys	Gln	Met 365	Tyr	Leu	Leu
Met 370	Thr	Lys	His	Leu	Leu	Asp 375	Gln	Gly	Phe	Val	Lys	Gly	Ser	Ala	Leu 380

SUBSTITUTE SHEET (RULE 26)

- 105 -

Val Thr Thr His His Trp Gly Tyr Ser Ser Arg Trp Trp Tyr Ile Ser
 385 390 395 400
 Thr Leu Leu Met Ser Asp Ala Leu Lys Glu Ala Asn Leu Gln Thr Gln
 405 410 415
 Val Tyr Asp Ser Leu Leu Trp Tyr Ser Arg Glu Phe Lys Ser Ser Phe
 420 425 430
 Asp Met Lys Val Ser Ala Asp Ser Ser Asp Leu Asp Tyr Phe Asn Thr
 435 440 445
 Leu Ser Arg Gln His Leu Ala Leu Leu Leu Leu Glu Pro Asp Asp Gln
 450 455 460
 Lys Arg Ile Asn Leu Val Asn Thr Phe Ser His Tyr Ile Thr Gly Ala
 465 470 475 480
 Leu Thr Gln Val Pro Pro Gly Gly Lys Asp Gly Leu Arg Pro Asp Gly
 485 490 495
 Thr Ala Trp Arg His Glu Gly Asn Tyr Pro Gly Tyr Ser Phe Pro Ala
 500 505 510
 Phe Lys Asn Ala Ser Gln Leu Ile Tyr Leu Leu Arg Asp Thr Pro Phe
 515 520 525
 Ser Val Gly Glu Ser Gly Trp Asn Asn Leu Lys Lys Ala Met Val Ser
 530 535 540
 Ala Trp Ile Tyr Ser Asn Pro Glu Val Gly Leu Pro Leu Ala Gly Arg
 545 550 555 560
 His Pro Phe Asn Ser Pro Ser Leu Lys Ser Val Ala Gln Gly Tyr Tyr
 565 570 575
 Trp Leu Ala Met Ser Ala Lys Ser Ser Pro Asp Lys Thr Leu Ala Ser
 580 585 590
 Ile Tyr Leu Ala Ile Ser Asp Lys Thr Gln Asn Glu Ser Thr Ala Ile
 595 600 605
 Phe Gly Glu Thr Ile Thr Pro Ala Ser Leu Pro Gln Gly Phe Tyr Ala
 610 615 620
 Phe Asn Gly Gly Ala Phe Gly Ile His Arg Trp Gln Asp Lys Met Val
 625 630 635 640
 Thr Leu Lys Ala Tyr Asn Thr Asn Val Trp Ser Ser Glu Ile Tyr Asn
 645 650 655
 Lys Asp Asn Arg Tyr Gly Arg Tyr Gln Ser His Gly Val Ala Gln Ile
 660 665 670
 Val Ser Asn Gly Ser Gln Leu Ser Gln Gly Tyr Gln Gln Glu Gly Trp
 675 680 685
 Asp Trp Asn Arg Met Gln Gly Ala Thr Thr Ile His Leu Pro Leu Lys
 690 695 700
 Asp Leu Asp Ser Pro Lys Pro His Thr Leu Met Gln Arg Gly Glu Arg
 705 710 715 720

SUBSTITUTE SHEET (RULE 26)

- 106 -

Gly Phe Ser Gly Thr Ser Ser Leu Glu Gly Gln Tyr Gly Met Met Ala
 725 730 735
 Phe Asp Leu Ile Tyr Pro Ala Asn Leu Glu Arg Phe Asp Pro Asn Phe
 740 745 750
 Thr Ala Lys Lys Ser Val Leu Ala Ala Asp Asn His Leu Ile Phe Ile
 755 760 765
 Gly Ser Asn Ile Asn Ser Ser Asp Lys Asn Lys Asn Val Glu Thr Thr
 770 775 780
 Leu Phe Gln His Ala Ile Thr Pro Thr Leu Asn Thr Leu Trp Ile Asn
 785 790 795 800
 Gly Gln Lys Ile Glu Asn Met Pro Tyr Gln Thr Thr Leu Gln Gln Gly
 805 810 815
 Asp Trp Leu Ile Asp Ser Asn Gly Asn Gly Tyr Leu Ile Thr Gln Ala
 820 825 830
 Glu Lys Val Asn Val Ser Arg Gln His Gln Val Ser Ala Glu Asn Lys
 835 840 845
 Asn Arg Gln Pro Thr Glu Gly Asn Phe Ser Ser Ala Trp Ile Asp His
 850 855 860
 Ser Thr Arg Pro Lys Asp Ala Ser Tyr Glu Tyr Met Val Phe Leu Asp
 865 870 875 880
 Ala Thr Pro Glu Lys Met Gly Glu Met Ala Gln Lys Phe Arg Glu Asn
 885 890 895
 Asn Gly Leu Tyr Gln Val Leu Arg Lys Asp Lys Asp Val His Ile Ile
 900 905 910
 Leu Asp Lys Leu Ser Asn Val Thr Gly Tyr Ala Phe Tyr Gln Pro Ala
 915 920 925
 Ser Ile Glu Asp Lys Trp Ile Lys Lys Val Asn Lys Pro Ala Ile Val
 930 935 940
 Met Thr His Arg Gln Lys Asp Thr Leu Ile Val Ser Ala Val Thr Pro
 945 950 955 960
 Asp Leu Asn Met Thr Arg Gln Lys Ala Ala Thr Pro Val Thr Ile Asn
 965 970 975
 Val Thr Ile Asn Gly Lys Trp Gln Ser Ala Asp Lys Asn Ser Glu Val
 980 985 990
 Lys Tyr Gln Val Ser Gly Asp Asn Thr Glu Leu Thr Phe Thr Ser Tyr
 995 1000 1005
 Phe Gly Ile Pro Gln Glu Ile Lys Leu Ser Pro Leu Pro
 1010 1015 1020

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3980 base pairs
 (B) TYPE: nucleic acid

SUBSTITUTE SHEET (RULE 26)

- 107 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGAATTCCAT CACTCAATCA TTAAATTTAG GCACAACGAT GGGCTATCAG CGTTATGACA	60
AATTTAATGA AGGACGCATT GGTTCCTACTG TTAGCCAGCG TTTCTAAGGA GAAAACATAT	120
GCCGATATTT CGTTTTACTG CACTTGCAAT GACATTGGGG CTATTATCAG CGCCTTATAA	180
CGCGATGGCA GCCACCAGCA ATCCTGCATT TGATCCTAAA AATCTGATGC AGTCAGAAAT	240
TTACCATTTT GCACAAAATA ACCCATTAGC AGACTTCTCA TCAGATAAAA ACTCAATACT	300
AACGTTATCT GATAAACGTA GCATTATGGG AAACCAATCT CTTTTATGGA AATGGAAAGG	360
TGGTAGTAGC TTTACTTTAC ATAAAAAACT GATTGTCCCC ACCGATAAAG AAGCATCTAA	420
AGCATGGGGA CGCTCATCTA CCCCCGTTTT CTCATTTTGG CTTTACAATG AAAAACCGAT	480
TGATGGTTAT CTTACTATCG ATTTCTGGAGA AAAACTCATT TCAACCAGTG AGGCTCAGGC	540
AGGCTTTAAA GTAAAATTAG ATTTCACTGG CTGGCGTGCT GTGGGAGTCT CTTTAAATAA	600
CGATCTTGAA AATCGAGAGA TGACCTTAAA TGCAACCAAT ACCTCCTCTG ATGGTACTCA	660
AGACAGCATT GGGCGTTCTT TAGGTGCTAA AGTCGATAGT ATTCGTTTTA AAGCGCCTTC	720
TAATGTGAGT CAGGGTGAAA TCTATATCGA CCGTATTATG TTTTCTGTCTG ATGATGCTCG	780
CTACCAATGG TCTGATTATC AAGTAAAAAC TCGCTTATCA GAACCTGAAA TTCAATTTCA	840
CAACGTAAAG CCACAACCTAC CTGTAACACC TGAAAATTTA GCGGCCATTG ATCTTATTCTG	900
CCAACGTCTA ATTAATGAAT TTGTCGGAGG TGAAAAAGAG ACAAACCTCG CATTAGAAGA	960
GAATATCAGC AAATTAAAAA GTGATTTCGA TGCTCTTAAT ATTCACACTT TAGCAAATGG	1020
TGGAACGCAA GGCAGACATC TGATCACTGA TAAACAAATC ATTATTTATC AACCAGAGAA	1080
TCTTAACTCC CAAGATAAAC AACTATTTGA TAATTATGTT ATTTTAGGTA ATTACACGAC	1140
ATTAATGTTT AATATTAGCC GTGCTTATGT GCTGGAAAAA GATCCACAC AAAAGGCGCA	1200
ACTAAAGCAG ATGTACTTAT TAATGACAAA GCATTTATTA GATCAAGGCT TTGTTAAAGG	1260
GAGTGCTTTA GTGACAACCC ATCACTGGGG ATACAGTTCT CGTTGGTGGT ATATTTCCAC	1320
GTTATTAATG TCTGATGCAC TAAAAGAAGC GAACCTACAA ACTCAAGTTT ATGATTTCATT	1380
ACTGTGGTAT TCACGTGAGT TTAAAAGTAG TTTTGATATG AAAGTAAGTG CTGATAGCTC	1440
TGATCTAGAT TATTTCAATA CCTTATCTCG CCAACATTTA GCCTTATTAT TACTAGAGCC	1500

SUBSTITUTE SHEET (RULE 26)

- 108 -

TGATGATCAA	AAGCGTATCA	ACTTAGTTAA	TACTTTCAGC	CATTATATCA	CTGGCGCATT	1560
AACGCAAGTG	CCACCGGGTG	GTAAAGATGG	TTTACGCCCT	GATGGTACAG	CATGGCGACA	1620
TGAAGGCAAC	TATCCGGGCT	ACTCTTTCCC	AGCCTTTAAA	AATGCCTCTC	AGCTTATTTA	1680
TTTATTACGC	GATACACCAT	TTTCAGTGGG	TGAAAGTGGT	TGGAATAACC	TGAAAAAAGC	1740
GATGGTTTCA	GCGTGGATCT	ACAGTAATCC	AGAAGTTGGA	TTACCGCTTG	CAGGAAGACA	1800
CCCTTTTAAC	TCACCTTCGT	TAAAATCAGT	CGCTCAAGGC	TATTACTGGC	TTGCCATGTC	1860
TGCAAAATCA	TCGCCTGATA	AAACACTTGC	ATCTATTTAT	CTTGCGATTA	GTGATAAAAC	1920
ACAAAATGAA	TCAACTGCTA	TTTTTGGAGA	AACTATTACA	CCAGCGTCTT	TACCTCAAGG	1980
TTTCTATGCC	TTTAATGGCG	GTGCTTTTGG	TATTCATCGT	TGGCAAGATA	AAATGGTGAC	2040
ACTGAAAGCT	TATAACACCA	ATGTTTGGTC	ATCTGAAATT	TATAACAAAG	ATAACCGTTA	2100
TGGCCGTTAC	CAAAGTCATG	GTGTCGCTCA	AATAGTGAGT	AATGGCTCGC	AGCTTTCACA	2160
GGGCTATCAG	CAAGAAGGTT	GGGATTGGAA	TAGAATGCAA	GGGGCAACCA	CTATTCACCT	2220
TCCTCTTAAA	GACTTAGACA	GTCCTAAACC	TCATACCTTA	ATGCAACGTG	GAGAGCGTGG	2280
ATTTAGCGGA	ACATCATCCC	TTGAAGGTCA	ATATGGCATG	ATGGCATTCT	ATCTTATTTA	2340
TCCCGCCAAT	CTTGAGCGTT	TTGATCCTAA	TTTCACTGCG	AAAAAGAGTG	TATTAGCCGC	2400
TGATAATCAC	TTAATTTTTA	TTGGTAGCAA	TATAAATAGT	AGTGATAAAA	ATAAAAAATGT	2460
TGAAACGACC	TTATTCCAAC	ATGCCATTAC	TCCAACATTA	AATACCCTTT	GGATTAATGG	2520
ACAAAAGATA	GAAAACATGC	CTTATCAAAC	AACACTTCAA	CAAGGTGATT	GGTTAATTGA	2580
TAGCAATGGC	AATGGTTACT	TAATTACTCA	AGCAGAAAAA	GTAAATGTAA	GTCGCCAACA	2640
TCAGGTTTCA	GCGGAAAATA	AAAATCGCCA	ACCGACAGAA	GGAAACTTTA	GCTCGGCATG	2700
GATCGATCAC	AGCACTCGCC	CCAAAGATGC	CAGTTATGAG	TATATGGTCT	TTTTAGATGC	2760
GACACCTGAA	AAAATGGGAG	AGATGGCACA	AAAATTCCGT	GAAAATAATG	GGTTATATCA	2820
GGTTCTTCGT	AAGGATAAAG	ACGTTCATAT	TATTCTCGAT	AAACTCAGCA	ATGTAACGGG	2880
ATATGCCCTTT	TATCAGCCAG	CATCAATTGA	AGACAAATGG	ATCAAAAAGG	TTAATAAACC	2940
TGCAATTGTG	ATGACTCATC	GACAAAAAGA	CACTCTTATT	GTCAGTGCAG	TTACACCTGA	3000
TTTAAATATG	ACTCGCCAAA	AAGCAGCAAC	TCCTGTCACC	ATCAATGTCA	CGATTAATGG	3060
CAAAATGGCA	TCTGCTGATA	AAAATAGTGA	AGTGAAATAT	CAGGTTTCTG	GTGATAACAC	3120
TGAACTGACG	TTTACGAGTT	ACTTTGGTAT	TCCACAAGAA	ATCAAACCTCT	CGCCACTCCC	3180
TTGATTTAAT	CAAAAGAACG	CTCTTGCGTT	CCTTTTTTTAT	TTGCAGGAAA	TCTGATTATG	3240
CTAATAAAAA	ACCCTTTAGC	CCACGCGGTT	ACATTAAGCC	TCTGTTTATC	ATTACCCGCA	3300
CAAGCATTAC	CCACTCTGTC	TCATGAAGCT	TTGCGCGATA	TTTATCTTTT	TGAAGGTGAA	3360

SUBSTITUTE SHEET (RULE 26)

- 109 -

```

TTACCCAATA CCCTTACCAC TTCAAATAAT AATCAATTAT CGCTAAGCAA ACAGCATGCT      3420
AAAGATGGTG AACAACTACT CAAATGGCAA TATCAACCAC AAGCAACATT AACACTAAAT      3480
AATATTGTTA ATTACCAAGA TGATAAAAAAT ACAGCCACAC CACTCACTTT TATGATGTGG      3540
ATTTATAATG AAAAACCTCA ATCTTCCCCA TTAACGTTAG CATTTAAACA AAATAATAAA      3600
ATTGCACTAA GTTTTAATGC TGAAGTTAAT TTTACGGGGT GGCGAGGTAT TGCTGTTCCCT      3660
TTTCGTGATA TGCAAGGCTC TGCGACAGGT CAACTTGATC AATTAGTGAT CACCGCTCCA      3720
AACCAAGCCG GAACACTCTT TTTTGATCAA ATCATCATGA GTGTACCGTT AGACAATCGT      3780
TGGGCAGTAC CTGACTATCA AACACCTTAC GTAAATAACG CAGTAAACAC GATGGTTAGT      3840
AAAAACTGGA GTGCATTATT GATGTACGAT CAGATGTTTC AAGCCCATTA CCCTACTTTA      3900
AACTTCGATA CTGAATTTTCG CGATGACCAA ACAGAAATGG CTTTCGATTTA TCAGCGCTTT      3960
GAATATTATC AAGGAATTCC                                     3980

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3980 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 188..3181

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

GGAATTCAT CACTCAATCA TTAAATTTAG GCACAACGAT GGGCTATCAG CGTTATGACA      60
AATTTAATGA AGGACGCATT GGTTCCTACTG TTAGCCAGCG TTTCTAAGGA GAAAAATAAT      120
GCCGATATTT CGTTTTACTG CACTTGCAAT GACATTGGGG CTATTATCAG CGCCTTATAA      180
CGCGGAT ATG GCC ACC AGC AAT CCT GCA TTT GAT CCT AAA AAT CTG ATG      229
      Met Ala Thr Ser Asn Pro Ala Phe Asp Pro Lys Asn Leu Met
          1              5              10
CAG TCA GAA ATT TAC CAT TTT GCA CAA AAT AAC CCA TTA GCA GAC TTC      277
Gln Ser Glu Ile Tyr His Phe Ala Gln Asn Asn Pro Leu Ala Asp Phe
      15              20              25              30
TCA TCA GAT AAA AAC TCA ATA CTA ACG TTA TCT GAT AAA CGT AGC ATT      325
Ser Ser Asp Lys Asn Ser Ile Leu Thr Leu Ser Asp Lys Arg Ser Ile
          35              40              45

```

SUBSTITUTE SHEET (RULE 26)

- 110 -

ATG GGA AAC CAA TCT CTT TTA TGG AAA TGG AAA GGT GGT AGT AGC TTT	373
Met Gly Asn Gln Ser Leu Leu Trp Lys Trp Lys Gly Gly Ser Ser Phe	
50 55 60	
ACT TTA CAT AAA AAA CTG ATT GTC CCC ACC GAT AAA GAA GCA TCT AAA	421
Thr Leu His Lys Lys Leu Ile Val Pro Thr Asp Lys Glu Ala Ser Lys	
65 70 75	
GCA TGG GGA CGC TCA TCT ACC CCC GTT TTC TCA TTT TGG CTT TAC AAT	469
Ala Trp Gly Arg Ser Ser Thr Pro Val Phe Ser Phe Trp Leu Tyr Asn	
80 85 90	
GAA AAA CCG ATT GAT GGT TAT CTT ACT ATC GAT TTC GGA GAA AAA CTC	517
Glu Lys Pro Ile Asp Gly Tyr Leu Thr Ile Asp Phe Gly Glu Lys Leu	
95 100 105 110	
ATT TCA ACC AGT GAG GCT CAG GCA GGC TTT AAA GTA AAA TTA GAT TTC	565
Ile Ser Thr Ser Glu Ala Gln Ala Gly Phe Lys Val Lys Leu Asp Phe	
115 120 125	
ACT GGC TGG CGT GCT GTG GGA GTC TCT TTA AAT AAC GAT CTT GAA AAT	613
Thr Gly Trp Arg Ala Val Gly Val Ser Leu Asn Asn Asp Leu Glu Asn	
130 135 140	
CGA GAG ATG ACC TTA AAT GCA ACC AAT ACC TCC TCT GAT GGT ACT CAA	661
Arg Glu Met Thr Leu Asn Ala Thr Asn Thr Ser Ser Asp Gly Thr Gln	
145 150 155	
GAC AGC ATT GGG CGT TCT TTA GGT GCT AAA GTC GAT AGT ATT CGT TTT	709
Asp Ser Ile Gly Arg Ser Leu Gly Ala Lys Val Asp Ser Ile Arg Phe	
160 165 170	
AAA GCG CCT TCT AAT GTG AGT CAG GGT GAA ATC TAT ATC GAC CGT ATT	757
Lys Ala Pro Ser Asn Val Ser Gln Gly Glu Ile Tyr Ile Asp Arg Ile	
175 180 185 190	
ATG TTT TCT GTC GAT GAT GCT CGC TAC CAA TGG TCT GAT TAT CAA GTA	805
Met Phe Ser Val Asp Asp Ala Arg Tyr Gln Trp Ser Asp Tyr Gln Val	
195 200 205	
AAA ACT CGC TTA TCA GAA CCT GAA ATT CAA TTT CAC AAC GTA AAG CCA	853
Lys Thr Arg Leu Ser Glu Pro Glu Ile Gln Phe His Asn Val Lys Pro	
210 215 220	
CAA CTA CCT GTA ACA CCT GAA AAT TTA GCG GCC ATT GAT CTT ATT CGC	901
Gln Leu Pro Val Thr Pro Glu Asn Leu Ala Ala Ile Asp Leu Ile Arg	
225 230 235	
CAA CGT CTA ATT AAT GAA TTT GTC GGA GGT GAA AAA GAG ACA AAC CTC	949
Gln Arg Leu Ile Asn Glu Phe Val Gly Gly Glu Lys Glu Thr Asn Leu	
240 245 250	
GCA TTA GAA GAG AAT ATC AGC AAA TTA AAA AGT GAT TTC GAT GCT CTT	997
Ala Leu Glu Glu Asn Ile Ser Lys Leu Lys Ser Asp Phe Asp Ala Leu	
255 260 265 270	
AAT ATT CAC ACT TTA GCA AAT GGT GGA ACG CAA GGC AGA CAT CTG ATC	1045
Asn Ile His Thr Leu Ala Asn Gly Gly Thr Gln Gly Arg His Leu Ile	
275 280 285	
ACT GAT AAA CAA ATC ATT ATT TAT CAA CCA GAG AAT CTT AAC TCC CAA	1093
Thr Asp Lys Gln Ile Ile Ile Tyr Gln Pro Glu Asn Leu Asn Ser Gln	

SUBSTITUTE SHEET (RULE 26)

- 111 -

290					295					300						
GAT	AAA	CAA	CTA	TTT	GAT	AAT	TAT	GTT	ATT	TTA	GGT	AAT	TAC	ACG	ACA	1141
Asp	Lys	Gln	Leu	Phe	Asp	Asn	Tyr	Val	Ile	Leu	Gly	Asn	Tyr	Thr	Thr	
		305					310					315				
TTA	ATG	TTT	AAT	ATT	AGC	CGT	GCT	TAT	GTG	CTG	GAA	AAA	GAT	CCC	ACA	1189
Leu	Met	Phe	Asn	Ile	Ser	Arg	Ala	Tyr	Val	Leu	Glu	Lys	Asp	Pro	Thr	
	320					325					330					
CAA	AAG	GCG	CAA	CTA	AAG	CAG	ATG	TAC	TTA	TTA	ATG	ACA	AAG	CAT	TTA	1237
Gln	Lys	Ala	Gln	Leu	Lys	Gln	Met	Tyr	Leu	Leu	Met	Thr	Lys	His	Leu	
335					340					345					350	
TTA	GAT	CAA	GGC	TTT	GTT	AAA	GGG	AGT	GCT	TTA	GTG	ACA	ACC	CAT	CAC	1285
Leu	Asp	Gln	Gly	Phe	Val	Lys	Gly	Ser	Ala	Leu	Val	Thr	Thr	His	His	
				355					360					365		
TGG	GGA	TAC	AGT	TCT	CGT	TGG	TGG	TAT	ATT	TCC	ACG	TTA	TTA	ATG	TCT	1333
Trp	Gly	Tyr	Ser	Ser	Arg	Trp	Trp	Tyr	Ile	Ser	Thr	Leu	Leu	Met	Ser	
			370					375					380			
GAT	GCA	CTA	AAA	GAA	GCG	AAC	CTA	CAA	ACT	CAA	GTT	TAT	GAT	TCA	TTA	1381
Asp	Ala	Leu	Lys	Glu	Ala	Asn	Leu	Gln	Thr	Gln	Val	Tyr	Asp	Ser	Leu	
		385					390					395				
CTG	TGG	TAT	TCA	CGT	GAG	TTT	AAA	AGT	AGT	TTT	GAT	ATG	AAA	GTA	AGT	1429
Leu	Trp	Tyr	Ser	Arg	Glu	Phe	Lys	Ser	Ser	Phe	Asp	Met	Lys	Val	Ser	
	400					405					410					
GCT	GAT	AGC	TCT	GAT	CTA	GAT	TAT	TTC	AAT	ACC	TTA	TCT	CGC	CAA	CAT	1477
Ala	Asp	Ser	Ser	Asp	Leu	Asp	Tyr	Phe	Asn	Thr	Leu	Ser	Arg	Gln	His	
415					420					425					430	
TTA	GCC	TTA	TTA	TTA	CTA	GAG	CCT	GAT	GAT	CAA	AAG	CGT	ATC	AAC	TTA	1525
Leu	Ala	Leu	Leu	Leu	Leu	Glu	Pro	Asp	Asp	Gln	Lys	Arg	Ile	Asn	Leu	
				435					440					445		
GTT	AAT	ACT	TTC	AGC	CAT	TAT	ATC	ACT	GGC	GCA	TTA	ACG	CAA	GTG	CCA	1573
Val	Asn	Thr	Phe	Ser	His	Tyr	Ile	Thr	Gly	Ala	Leu	Thr	Gln	Val	Pro	
			450					455					460			
CCG	GGT	GGT	AAA	GAT	GGT	TTA	CGC	CCT	GAT	GGT	ACA	GCA	TGG	CGA	CAT	1621
Pro	Gly	Gly	Lys	Asp	Gly	Leu	Arg	Pro	Asp	Gly	Thr	Ala	Trp	Arg	His	
		465					470					475				
GAA	GGC	AAC	TAT	CCG	GGC	TAC	TCT	TTC	CCA	GCC	TTT	AAA	AAT	GCC	TCT	1669
Glu	Gly	Asn	Tyr	Pro	Gly	Tyr	Ser	Phe	Pro	Ala	Phe	Lys	Asn	Ala	Ser	
	480					485					490					
CAG	CTT	ATT	TAT	TTA	TTA	CGC	GAT	ACA	CCA	TTT	TCA	GTG	GGT	GAA	AGT	1717
Gln	Leu	Ile	Tyr	Leu	Leu	Arg	Asp	Thr	Pro	Phe	Ser	Val	Gly	Glu	Ser	
495					500					505				510		
GGT	TGG	AAT	AAC	CTG	AAA	AAA	GCG	ATG	GTT	TCA	GCG	TGG	ATC	TAC	AGT	1765
Gly	Trp	Asn	Asn	Leu	Lys	Lys	Ala	Met	Val	Ser	Ala	Trp	Ile	Tyr	Ser	
				515					520					525		
AAT	CCA	GAA	GTT	GGA	TTA	CCG	CTT	GCA	GGA	AGA	CAC	CCT	TTT	AAC	TCA	1813
Asn	Pro	Glu	Val	Gly	Leu	Pro	Leu	Ala	Gly	Arg	His	Pro	Phe	Asn	Ser	
			530					535					540			

SUBSTITUTE SHEET (RULE 26)

- 112 -

CCT	TCG	TTA	AAA	TCA	GTC	GCT	CAA	GGC	TAT	TAC	TGG	CTT	GCC	ATG	TCT	1861
Pro	Ser	Leu	Lys	Ser	Val	Ala	Gln	Gly	Tyr	Tyr	Trp	Leu	Ala	Met	Ser	
		545					550					555				
GCA	AAA	TCA	TCG	CCT	GAT	AAA	ACA	CTT	GCA	TCT	ATT	TAT	CTT	GCG	ATT	1909
Ala	Lys	Ser	Ser	Pro	Asp	Lys	Thr	Leu	Ala	Ser	Ile	Tyr	Leu	Ala	Ile	
	560					565					570					
AGT	GAT	AAA	ACA	CAA	AAT	GAA	TCA	ACT	GCT	ATT	TTT	GGA	GAA	ACT	ATT	1957
Ser	Asp	Lys	Thr	Gln	Asn	Glu	Ser	Thr	Ala	Ile	Phe	Gly	Glu	Thr	Ile	
575					580					585					590	
ACA	CCA	GCG	TCT	TTA	CCT	CAA	GGT	TTC	TAT	GCC	TTT	AAT	GGC	GGT	GCT	2005
Thr	Pro	Ala	Ser	Leu	Pro	Gln	Gly	Phe	Tyr	Ala	Phe	Asn	Gly	Gly	Ala	
				595					600					605		
TTT	GGT	ATT	CAT	CGT	TGG	CAA	GAT	AAA	ATG	GTG	ACA	CTG	AAA	GCT	TAT	2053
Phe	Gly	Ile	His	Arg	Trp	Gln	Asp	Lys	Met	Val	Thr	Leu	Lys	Ala	Tyr	
			610					615					620			
AAC	ACC	AAT	GTT	TGG	TCA	TCT	GAA	ATT	TAT	AAC	AAA	GAT	AAC	CGT	TAT	2101
Asn	Thr	Asn	Val	Trp	Ser	Ser	Glu	Ile	Tyr	Asn	Lys	Asp	Asn	Arg	Tyr	
		625					630					635				
GGC	CGT	TAC	CAA	AGT	CAT	GGT	GTC	GCT	CAA	ATA	GTG	AGT	AAT	GGC	TCG	2149
Gly	Arg	Tyr	Gln	Ser	His	Gly	Val	Ala	Gln	Ile	Val	Ser	Asn	Gly	Ser	
	640					645					650					
CAG	CTT	TCA	CAG	GGC	TAT	CAG	CAA	GAA	GGT	TGG	GAT	TGG	AAT	AGA	ATG	2197
Gln	Leu	Ser	Gln	Gly	Tyr	Gln	Gln	Glu	Gly	Trp	Asp	Trp	Asn	Arg	Met	
655					660					665					670	
CAA	GGG	GCA	ACC	ACT	ATT	CAC	CTT	CCT	CTT	AAA	GAC	TTA	GAC	AGT	CCT	2245
Gln	Gly	Ala	Thr	Thr	Ile	His	Leu	Pro	Leu	Lys	Asp	Leu	Asp	Ser	Pro	
				675					680					685		
AAA	CCT	CAT	ACC	TTA	ATG	CAA	CGT	GGA	GAG	CGT	GGA	TTT	AGC	GGA	ACA	2293
Lys	Pro	His	Thr	Leu	Met	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Ser	Gly	Thr	
			690					695					700			
TCA	TCC	CTT	GAA	GGT	CAA	TAT	GGC	ATG	ATG	GCA	TTC	GAT	CTT	ATT	TAT	2341
Ser	Ser	Leu	Glu	Gly	Gln	Tyr	Gly	Met	Met	Ala	Phe	Asp	Leu	Ile	Tyr	
		705					710					715				
CCC	GCC	AAT	CTT	GAG	CGT	TTT	GAT	CCT	AAT	TTC	ACT	GCG	AAA	AAG	AGT	2389
Pro	Ala	Asn	Leu	Glu	Arg	Phe	Asp	Pro	Asn	Phe	Thr	Ala	Lys	Lys	Ser	
	720					725					730					
GTA	TTA	GCC	GCT	GAT	AAT	CAC	TTA	ATT	TTT	ATT	GGT	AGC	AAT	ATA	AAT	2437
Val	Leu	Ala	Ala	Asp	Asn	His	Leu	Ile	Phe	Ile	Gly	Ser	Asn	Ile	Asn	
735					740					745					750	
AGT	AGT	GAT	AAA	AAT	AAA	AAT	GTT	GAA	ACG	ACC	TTA	TTC	CAA	CAT	GCC	2485
Ser	Ser	Asp	Lys	Asn	Lys	Asn	Val	Glu	Thr	Thr	Leu	Phe	Gln	His	Ala	
				755					760					765		
ATT	ACT	CCA	ACA	TTA	AAT	ACC	CTT	TGG	ATT	AAT	GGA	CAA	AAG	ATA	GAA	2533
Ile	Thr	Pro	Thr	Leu	Asn	Thr	Leu	Trp	Ile	Asn	Gly	Gln	Lys	Ile	Glu	
			770					775					780			
AAC	ATG	CCT	TAT	CAA	ACA	ACA	CTT	CAA	CAA	GGT	GAT	TGG	TTA	ATT	GAT	2581
Asn	Met	Pro	Tyr	Gln	Thr	Thr	Leu	Gln	Gln	Gly	Asp	Trp	Leu	Ile	Asp	

SUBSTITUTE SHEET (RULE 26)

- 113 -

785	790	795	
AGC AAT GGC AAT GGT TAC TTA ATT ACT CAA GCA GAA AAA GTA AAT GTA Ser Asn Gly Asn Gly Tyr Leu Ile Thr Gln Ala Glu Lys Val Asn Val 800 805 810			2629
AGT CGC CAA CAT CAG GTT TCA GCG GAA AAT AAA AAT CGC CAA CCG ACA Ser Arg Gln His Gln Val Ser Ala Glu Asn Lys Asn Arg Gln Pro Thr 815 820 825 830			2677
GAA GGA AAC TTT AGC TCG GCA TGG ATC GAT CAC AGC ACT CGC CCC AAA Glu Gly Asn Phe Ser Ser Ala Trp Ile Asp His Ser Thr Arg Pro Lys 835 840 845			2725
GAT GCC AGT TAT GAG TAT ATG GTC TTT TTA GAT GCG ACA CCT GAA AAA Asp Ala Ser Tyr Glu Tyr Met Val Phe Leu Asp Ala Thr Pro Glu Lys 850 855 860			2773
ATG GGA GAG ATG GCA CAA AAA TTC CGT GAA AAT AAT GGG TTA TAT CAG Met Gly Glu Met Ala Gln Lys Phe Arg Glu Asn Asn Gly Leu Tyr Gln 865 870 875			2821
GTT CTT CGT AAG GAT AAA GAC GTT CAT ATT ATT CTC GAT AAA CTC AGC Val Leu Arg Lys Asp Lys Asp Val His Ile Ile Leu Asp Lys Leu Ser 880 885 890			2869
AAT GTA ACG GGA TAT GCC TTT TAT CAG CCA GCA TCA ATT GAA GAC AAA Asn Val Thr Gly Tyr Ala Phe Tyr Gln Pro Ala Ser Ile Glu Asp Lys 895 900 905 910			2917
TGG ATC AAA AAG GTT AAT AAA CCT GCA ATT GTG ATG ACT CAT CGA CAA Trp Ile Lys Lys Val Asn Lys Pro Ala Ile Val Met Thr His Arg Gln 915 920 925			2965
AAA GAC ACT CTT ATT GTC AGT GCA GTT ACA CCT GAT TTA AAT ATG ACT Lys Asp Thr Leu Ile Val Ser Ala Val Thr Pro Asp Leu Asn Met Thr 930 935 940			3013
CGC CAA AAA GCA GCA ACT CCT GTC ACC ATC AAT GTC ACG ATT AAT GGC Arg Gln Lys Ala Ala Thr Pro Val Thr Ile Asn Val Thr Ile Asn Gly 945 950 955			3061
AAA TGG CAA TCT GCT GAT AAA AAT AGT GAA GTG AAA TAT CAG GTT TCT Lys Trp Gln Ser Ala Asp Lys Asn Ser Glu Val Lys Tyr Gln Val Ser 960 965 970			3109
GGT GAT AAC ACT GAA CTG ACG TTT ACG AGT TAC TTT GGT ATT CCA CAA Gly Asp Asn Thr Glu Leu Thr Phe Thr Ser Tyr Phe Gly Ile Pro Gln 975 980 985 990			3157
GAA ATC AAA CTC TCG CCA CTC CCT TGATTTAATC AAAAGAACGC TCTTGC GTTC Glu Ile Lys Leu Ser Pro Leu Pro 995			3211
CTTTTTTATT TGCAGGAAAT CTGATTATGC TAATAAAAAA CCCTTTAGCC CACGCGGTTA			3271
CATTAAGCCT CTGTTTATCA TTACCCGCAC AAGCATTACC CACTCTGTCT CATGAAGCTT			3331
TCGGCGATAT TTATCTTTTT GAAGGTGAAT TACCCAATAC CCTTACCACT TCAATAATA			3391
ATCAATTATC GCTAAGCAAA CAGCATGCTA AAGATGGTGA ACAATCACTC AAATGGCAAT			3451

SUBSTITUTE SHEET (RULE 26)

- 114 -

ATCAACCACA AGCAACATTA AACTAAATA ATATTGTTAA TTACCAAGAT GATAAAAATA 3511
 CAGCCACACC ACTCACTTTT ATGATGTGGA TTTATAATGA AAAACCTCAA TCTTCCCCAT 3571
 TAACGTTAGC ATTTAAACAA AATAATAAAA TTGCACTAAG TTTTAATGCT GAACTTAATT 3631
 TTACGGGGTG GCGAGGTATT GCTGTTCTT TTCGTGATAT GCAAGGCTCT GCGACAGGTC 3691
 AACTTGATCA ATTAGTGATC ACCGCTCCAA ACCAAGCCGG AACACTCTTT TTTGATCAAA 3751
 TCATCATGAG TGTACCGTTA GACAATCGTT GGGCAGTACC TGACTATCAA ACACCTTACG 3811
 TAAATAACGC AGTAAACACG ATGGTTAGTA AAAACTGGAG TGCATTATTG ATGTACGATC 3871
 AGATGTTTCA AGCCCATAC CCTACTTTAA ACTTCGATAC TGAATTTTCG GATGACCAA 3931
 CAGAAATGGC TTCGATTTAT CAGCGCTTTG AATATTATCA AGGAATTCC 3980

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 998 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Ala	Thr	Ser	Asn	Pro	Ala	Phe	Asp	Pro	Lys	Asn	Leu	Met	Gln	Ser	1	5	10	15
Glu	Ile	Tyr	His	Phe	Ala	Gln	Asn	Asn	Pro	Leu	Ala	Asp	Phe	Ser	Ser	20	25	30	
Asp	Lys	Asn	Ser	Ile	Leu	Thr	Leu	Ser	Asp	Lys	Arg	Ser	Ile	Met	Gly	35	40	45	
Asn	Gln	Ser	Leu	Leu	Trp	Lys	Trp	Lys	Gly	Gly	Ser	Ser	Phe	Thr	Leu	50	55	60	
His	Lys	Lys	Leu	Ile	Val	Pro	Thr	Asp	Lys	Glu	Ala	Ser	Lys	Ala	Trp	65	70	75	80
Gly	Arg	Ser	Ser	Thr	Pro	Val	Phe	Ser	Phe	Trp	Leu	Tyr	Asn	Glu	Lys	85	90	95	
Pro	Ile	Asp	Gly	Tyr	Leu	Thr	Ile	Asp	Phe	Gly	Glu	Lys	Leu	Ile	Ser	100	105	110	
Thr	Ser	Glu	Ala	Gln	Ala	Gly	Phe	Lys	Val	Lys	Leu	Asp	Phe	Thr	Gly	115	120	125	
Trp	Arg	Ala	Val	Gly	Val	Ser	Leu	Asn	Asn	Asp	Leu	Glu	Asn	Arg	Glu	130	135	140	
Met	Thr	Leu	Asn	Ala	Thr	Asn	Thr	Ser	Ser	Asp	Gly	Thr	Gln	Asp	Ser	145	150	155	160
Ile	Gly	Arg	Ser	Leu	Gly	Ala	Lys	Val	Asp	Ser	Ile	Arg	Phe	Lys	Ala	165	170	175	

SUBSTITUTE SHEET (RULE 26)

- 115 -

Pro Ser Asn Val Ser Gln Gly Glu Ile Tyr Ile Asp Arg Ile Met Phe
 180 185 190
 Ser Val Asp Asp Ala Arg Tyr Gln Trp Ser Asp Tyr Gln Val Lys Thr
 195 200 205
 Arg Leu Ser Glu Pro Glu Ile Gln Phe His Asn Val Lys Pro Gln Leu
 210 215 220
 Pro Val Thr Pro Glu Asn Leu Ala Ala Ile Asp Leu Ile Arg Gln Arg
 225 230 235 240
 Leu Ile Asn Glu Phe Val Gly Gly Glu Lys Glu Thr Asn Leu Ala Leu
 245 250 255
 Glu Glu Asn Ile Ser Lys Leu Lys Ser Asp Phe Asp Ala Leu Asn Ile
 260 265 270
 His Thr Leu Ala Asn Gly Gly Thr Gln Gly Arg His Leu Ile Thr Asp
 275 280 285
 Lys Gln Ile Ile Ile Tyr Gln Pro Glu Asn Leu Asn Ser Gln Asp Lys
 290 295 300
 Gln Leu Phe Asp Asn Tyr Val Ile Leu Gly Asn Tyr Thr Thr Leu Met
 305 310 315 320
 Phe Asn Ile Ser Arg Ala Tyr Val Leu Glu Lys Asp Pro Thr Gln Lys
 325 330 335
 Ala Gln Leu Lys Gln Met Tyr Leu Leu Met Thr Lys His Leu Leu Asp
 340 345 350
 Gln Gly Phe Val Lys Gly Ser Ala Leu Val Thr Thr His His Trp Gly
 355 360 365
 Tyr Ser Ser Arg Trp Trp Tyr Ile Ser Thr Leu Leu Met Ser Asp Ala
 370 375 380
 Leu Lys Glu Ala Asn Leu Gln Thr Gln Val Tyr Asp Ser Leu Leu Trp
 385 390 395 400
 Tyr Ser Arg Glu Phe Lys Ser Ser Phe Asp Met Lys Val Ser Ala Asp
 405 410 415
 Ser Ser Asp Leu Asp Tyr Phe Asn Thr Leu Ser Arg Gln His Leu Ala
 420 425 430
 Leu Leu Leu Leu Glu Pro Asp Asp Gln Lys Arg Ile Asn Leu Val Asn
 435 440 445
 Thr Phe Ser His Tyr Ile Thr Gly Ala Leu Thr Gln Val Pro Pro Gly
 450 455 460
 Gly Lys Asp Gly Leu Arg Pro Asp Gly Thr Ala Trp Arg His Glu Gly
 465 470 475 480
 Asn Tyr Pro Gly Tyr Ser Phe Pro Ala Phe Lys Asn Ala Ser Gln Leu
 485 490 495
 Ile Tyr Leu Leu Arg Asp Thr Pro Phe Ser Val Gly Glu Ser Gly Trp
 500 505 510

SUBSTITUTE SHEET (RULE 26)

- 116 -

Asn Asn Leu Lys Lys Ala Met Val Ser Ala Trp Ile Tyr Ser Asn Pro
 515 520 525
 Glu Val Gly Leu Pro Leu Ala Gly Arg His Pro Phe Asn Ser Pro Ser
 530 535 540
 Leu Lys Ser Val Ala Gln Gly Tyr Tyr Trp Leu Ala Met Ser Ala Lys
 545 550 555 560
 Ser Ser Pro Asp Lys Thr Leu Ala Ser Ile Tyr Leu Ala Ile Ser Asp
 565 570 575
 Lys Thr Gln Asn Glu Ser Thr Ala Ile Phe Gly Glu Thr Ile Thr Pro
 580 585 590
 Ala Ser Leu Pro Gln Gly Phe Tyr Ala Phe Asn Gly Gly Ala Phe Gly
 595 600 605
 Ile His Arg Trp Gln Asp Lys Met Val Thr Leu Lys Ala Tyr Asn Thr
 610 615 620
 Asn Val Trp Ser Ser Glu Ile Tyr Asn Lys Asp Asn Arg Tyr Gly Arg
 625 630 635 640
 Tyr Gln Ser His Gly Val Ala Gln Ile Val Ser Asn Gly Ser Gln Leu
 645 650 655
 Ser Gln Gly Tyr Gln Gln Glu Gly Trp Asp Trp Asn Arg Met Gln Gly
 660 665 670
 Ala Thr Thr Ile His Leu Pro Leu Lys Asp Leu Asp Ser Pro Lys Pro
 675 680 685
 His Thr Leu Met Gln Arg Gly Glu Arg Gly Phe Ser Gly Thr Ser Ser
 690 695 700
 Leu Glu Gly Gln Tyr Gly Met Met Ala Phe Asp Leu Ile Tyr Pro Ala
 705 710 715 720
 Asn Leu Glu Arg Phe Asp Pro Asn Phe Thr Ala Lys Lys Ser Val Leu
 725 730 735
 Ala Ala Asp Asn His Leu Ile Phe Ile Gly Ser Asn Ile Asn Ser Ser
 740 745 750
 Asp Lys Asn Lys Asn Val Glu Thr Thr Leu Phe Gln His Ala Ile Thr
 755 760 765
 Pro Thr Leu Asn Thr Leu Trp Ile Asn Gly Gln Lys Ile Glu Asn Met
 770 775 780
 Pro Tyr Gln Thr Thr Leu Gln Gln Gly Asp Trp Leu Ile Asp Ser Asn
 785 790 795 800
 Gly Asn Gly Tyr Leu Ile Thr Gln Ala Glu Lys Val Asn Val Ser Arg
 805 810 815
 Gln His Gln Val Ser Ala Glu Asn Lys Asn Arg Gln Pro Thr Glu Gly
 820 825 830
 Asn Phe Ser Ser Ala Trp Ile Asp His Ser Thr Arg Pro Lys Asp Ala
 835 840 845

SUBSTITUTE SHEET (RULE 26)

- 117 -

Ser Tyr Glu Tyr Met Val Phe Leu Asp Ala Thr Pro Glu Lys Met Gly
 850 855 860
 Glu Met Ala Gln Lys Phe Arg Glu Asn Asn Gly Leu Tyr Gln Val Leu
 865 870 875
 Arg Lys Asp Lys Asp Val His Ile Ile Leu Asp Lys Leu Ser Asn Val
 885 890 895
 Thr Gly Tyr Ala Phe Tyr Gln Pro Ala Ser Ile Glu Asp Lys Trp Ile
 900 905 910
 Lys Lys Val Asn Lys Pro Ala Ile Val Met Thr His Arg Gln Lys Asp
 915 920 925
 Thr Leu Ile Val Ser Ala Val Thr Pro Asp Leu Asn Met Thr Arg Gln
 930 935 940
 Lys Ala Ala Thr Pro Val Thr Ile Asn Val Thr Ile Asn Gly Lys Trp
 945 950 955
 Gln Ser Ala Asp Lys Asn Ser Glu Val Lys Tyr Gln Val Ser Gly Asp
 965 970 975
 Asn Thr Glu Leu Thr Phe Thr Ser Tyr Phe Gly Ile Pro Gln Glu Ile
 980 985 990
 Lys Leu Ser Pro Leu Pro
 995

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CAYTTYGCNC ARAAYAAYCC N

21

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

- 118 -

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CACTTCGCNC AAAATAATCC

20

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CACTTCGCNC AAAACAACCC

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CACTTCGCNC AAAACAATCC

20

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

SUBSTITUTE SHEET (RULE 26)

- 119 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CACTTCGCNC AAAATAACCC

20

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CACTTCGCNC AGAATAATCC

20

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CACTTCGCNC AGAACAACCC

20

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

- 120 -

CACTTCGCNC AGAACAATCC

20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CACTTCGCNC AGAATAACCC

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GARGCNCARG CNGGNTTYAA R

21

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

YTTRAANCCN GCYTGNGCYT C

21

- 121 -

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGAARCCNG CYTGGGCTTC

20

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTGAARCCNG CYTGAGCTTC

20

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTGAARCCNG CYTGTGCTTC

20

(2) INFORMATION FOR SEQ ID NO:20:

SUBSTITUTE SHEET (RULE 26)

- 122 -

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: YES
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTGAARCCNG CYTGCGCTTC

20

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: YES
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTGAARCCNG CYTGGGCCTC

20

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: YES
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTGAARCCNG CYTGAGCCTC

20

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs

SUBSTITUTE SHEET (RULE 26)

- 123 -

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTGAARCCNG CYTGTGCCTC

20

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTGAARCCNG CYTGCGCCTC

20

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGNGCNAARG TNGAYTCN

18

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- 124 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGNGCNAARG TNGAYAGY

18

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

NGARTCNACY TTNGCNCC

18

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

RCTRTCACY TTNGCNCC

18

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- 125 -

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAGTCNACYT TRGCGCC

17

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAGTCNACYT TRGCACC

17

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GAGTCNACYT TRGCTCC

17

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

SUBSTITUTE SHEET (RULE 26)

- 126 -

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GAGTCNACYT TRGCCCC

17

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GAGTCNACYT TYGCGCC

17

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GAGTCNACYT TYGCACC

17

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

SUBSTITUTE SHEET (RULE 26)

- 127 -

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAGTCNACYT TYGCTCC

17

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GAGTCNACYT TYGCCCC

17

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCCAGCGTTT CTAAGGAGAA AACATATGCC GATATTTCGT TTTACTGC

48

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

- 128 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCGCCTTATA ACGCGCATAT GGCCACCAGC AATCCTG

37

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6519 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3238..6276

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGAATTCCAT CACTCAATCA TTAAATTTAG GCACAACGAT GGGCTATCAG CGTTATGACA	60
AATTTAATGA AGGACGCATT GGTTCACCTG TTAGCCAGCG TTTCTAAGGA GAAAAATAAT	120
GCCGATATTT CGTTTTACTG CACTTGCAAT GACATTGGGG CTATTATCAG CGCCTTATAA	180
CGCGATGGCA GCCACCAGCA ATCCTGCATT TGATCCTAAA AATCTGATGC AGTCAGAAAT	240
TTACCATTTT GCACAAAATA ACCCATTAGC AGACTTCTCA TCAGATAAAA ACTCAATACT	300
AACGTTATCT GATAAACGTA GCATTATGGG AAACCAATCT CTTTTATGGA AATGGAAAGG	360
TGGTAGTAGC TTTACTTTAC ATAAAAAACT GATTGTCCCC ACCGATAAAG AAGCATCTAA	420
AGCATGGGGA CGCTCATCTA CCCCCGTTTT CTCATTTTGG CTTTACAATG AAAAACCGAT	480
TGATGGTTAT CTTACTATCG ATTTCCGAGA AAAACTCATT TCAACCAAGT AGGCTCAGGC	540
AGGCTTTAAA GTAAAATTAG ATTTCACTGG CTGGCGTGCT GTGGGAGTCT CTTTAAATAA	600
CGATCTTGAA AATCGAGAGA TGACCTTAAA TGCAACCAAT ACCTCCTCTG ATGGTACTCA	660
AGACAGCATT GGGCGTTCTT TAGGTGCTAA AGTCGATAGT ATTCGTTTTA AAGCGCCTTC	720
TAATGTGAGT CAGGGTGAAA TCTATATCGA CCGTATTATG TTTTCTGTCG ATGATGCTCG	780
CTACCAATGG TCTGATTATC AAGTAAAAAC TCGCTTATCA GAACCTGAAA TTCAATTTCA	840
CAACGTAAAG CCACAACTAC CTGTAACACC TGAAAATTTA GCGGCCATTG ATCTTATTCTG	900
CCAACGTCTA ATTAATGAAT TTGTCGGAGG TGAAAAAGAG ACAAACCTCG CATTAGAAGA	960
GAATATCAGC AAATTAAAAA GTGATTTCTG TGCTCTTAAT ATTCACACTT TAGCAAATGG	1020
TGGAACGCAA GGCAGACATC TGATCACTGA TAAACAAATC ATTATTTATC AACCAGAGAA	1080

SUBSTITUTE SHEET (RULE 26)

- 129 -

TCTTAACTCC	CAAGATAAAC	AACTATTTGA	TAATTATGTT	ATTTTAGGTA	ATTACACGAC	1140
ATTAATGTTT	AATATTAGCC	GTGCTTATGT	GCTGGAAAAA	GATCCACAC	AAAAGGCGCA	1200
ACTAAAGCAG	ATGTACTTAT	TAATGACAAA	GCATTTATTA	GATCAAGGCT	TTGTTAAAGG	1260
GAGTGCTTTA	GTGACAACCC	ATCACTGGGG	ATACAGTTCT	CGTTGGTGGT	ATATTTCCAC	1320
GTTATTAATG	TCTGATGCAC	TAAAAGAAGC	GAACCTACAA	ACTCAAGTTT	ATGATTCATT	1380
ACTGTGGTAT	TCACGTGAGT	TTAAAAGTAG	TTTTGATATG	AAAGTAAGTG	CTGATAGCTC	1440
TGATCTAGAT	TATTTCAATA	CCTTATCTCG	CCAACATTTA	GCCTTATTAT	TACTAGAGCC	1500
TGATGATCAA	AAGCGTATCA	ACTTAGTTAA	TACTTTTCAGC	CATTATATCA	CTGGCGCATT	1560
AACGCAAGTG	CCACCGGGTG	GTAAAGATGG	TTTACGCCCT	GATGGTACAG	CATGGCGACA	1620
TGAAGGCAAC	TATCCGGGCT	ACTCTTTCCC	AGCCTTTAAA	AATGCCTCTC	AGCTTATTTA	1680
TTTATTACGC	GATACACCAT	TTTCAGTGGG	TGAAAGTGGT	TGGAATAACC	TGAAAAAAGC	1740
GATGGTTTCA	GCGTGGATCT	ACAGTAATCC	AGAAGTTGGA	TTACCGCTTG	CAGGAAGACA	1800
CCCTTTTAAC	TCACCTTCGT	TAAATCAGT	CGCTCAAGGC	TATTACTGGC	TTGCCATGTC	1860
TGCAAAATCA	TCGCCTGATA	AAACACTTGC	ATCTATTTAT	CTTGCGATTA	GTGATAAAAC	1920
ACAAATGAA	TCAACTGCTA	TTTTTGGAGA	AACTATTACA	CCAGCGTCTT	TACCTCAAGG	1980
TTTCTATGCC	TTTAATGGCG	GTGCTTTTGG	TATTCATCGT	TGGCAAGATA	AAATGGTGAC	2040
ACTGAAAGCT	TATAACACCA	ATGTTTGGTC	ATCTGAAATT	TATAACAAAG	ATAACCGTTA	2100
TGGCCGTTAC	CAAAGTCATG	GTGTCGCTCA	AATAGTGAGT	AATGGCTCGC	AGCTTTCACA	2160
GGGCTATCAG	CAAGAAGGTT	GGGATTGGAA	TAGAATGCAA	GGGGCAACCA	CTATTCACCT	2220
TCCTCTTAAA	GACTTAGACA	GTCCTAAACC	TCATACCTTA	ATGCAACGTG	GAGAGCGTGG	2280
ATTTAGCGGA	ACATCATCCC	TTGAAGGTCA	ATATGGCATG	ATGGCATTCT	ATCTTATTTA	2340
TCCCGCCAAT	CTTGAGCGTT	TTGATCCTAA	TTTCACTGCG	AAAAAGAGTG	TATTAGCCGC	2400
TGATAATCAC	TTAATTTTAA	TTGGTAGCAA	TATAAATAGT	AGTGATAAAA	ATAAAAATGT	2460
TGAAACGACC	TTATTCCAAC	ATGCCATTAC	TCCAACATTA	AATACCCTTT	GGATTAATGG	2520
ACAAAAGATA	GAAAACATGC	CTTATCAAAC	AACACTTCAA	CAAGGTGATT	GGTTAATTGA	2580
TAGCAATGGC	AATGGTTACT	TAATTACTCA	AGCAGAAAAA	GTAAATGTAA	GTCGCCAACA	2640
TCAGGTTTCA	GCGGAAAAATA	AAAATCGCCA	ACCGACAGAA	GGAAACTTTA	GCTCGGCATG	2700
GATCGATCAC	AGCACTCGCC	CCAAAGATGC	CAGTTATGAG	TATATGGTCT	TTTTAGATGC	2760
GACACCTGAA	AAAATGGGAG	AGATGGCACA	AAAATTCCGT	GAAAATAATG	GGTTATATCA	2820
GGTTCCTTCGT	AAGGATAAAG	ACGTTTATAT	TATTCTCGAT	AACTCAGCA	ATGTAACGGG	2880
ATATGCCTTT	TATCAGCCAG	CATCAATTGA	AGACAAATGG	ATCAAAAAGG	TTAATAAACC	2940

SUBSTITUTE SHEET (RULE 26)

- 130 -

TGCAATTGTG ATGACTCATC GACAAAAAGA CACTCTTATT GTCAGTGCAG TTACACCTGA	3000
TTTAAATATG ACTCGCCAAA AAGCAGCAAC TCCTGTCACC ATCAATGTCA CGATTAATGG	3060
CAAATGGCAA TCTGCTGATA AAAATAGTGA AGTGAAATAT CAGGTTTCTG GTGATAACAC	3120
TGAACTGACG TTTACGAGTT ACTTTGGTAT TCCACAAGAA ATCAAACCTCT CGCCACTCCC	3180
TTGATTTAAT CAAAAGAACG CTCTTGCGTT CCTTTTTTAT TTGCAGGAAA TCTGATT	3237
ATG CTA ATA AAA AAC CCT TTA GCC CAC GCG GTT ACA TTA AGC CTC TGT Met Leu Ile Lys Asn Pro Leu Ala His Ala Val Thr Leu Ser Leu Cys 1 5 10 15	3285
TTA TCA TTA CCC GCA CAA GCA TTA CCC ACT CTG TCT CAT GAA GCT TTC Leu Ser Leu Pro Ala Gln Ala Leu Pro Thr Leu Ser His Glu Ala Phe 20 25 30	3333
GGC GAT ATT TAT CTT TTT GAA GGT GAA TTA CCC AAT ACC CTT ACC ACT Gly Asp Ile Tyr Leu Phe Glu Gly Glu Leu Pro Asn Thr Leu Thr Thr 35 40 45	3381
TCA AAT AAT AAT CAA TTA TCG CTA AGC AAA CAG CAT GCT AAA GAT GGT Ser Asn Asn Asn Gln Leu Ser Leu Ser Lys Gln His Ala Lys Asp Gly 50 55 60	3429
GAA CAA TCA CTC AAA TGG CAA TAT CAA CCA CAA GCA ACA TTA ACA CTA Glu Gln Ser Leu Lys Trp Gln Tyr Gln Pro Gln Ala Thr Leu Thr Leu 65 70 75 80	3477
AAT AAT ATT GTT AAT TAC CAA GAT GAT AAA AAT ACA GCC ACA CCA CTC Asn Asn Ile Val Asn Tyr Gln Asp Asp Lys Asn Thr Ala Thr Pro Leu 85 90 95	3525
ACT TTT ATG ATG TGG ATT TAT AAT GAA AAA CCT CAA TCT TCC CCA TTA Thr Phe Met Met Trp Ile Tyr Asn Glu Lys Pro Gln Ser Ser Pro Leu 100 105 110	3573
ACG TTA GCA TTT AAA CAA AAT AAT AAA ATT GCA CTA AGT TTT AAT GCT Thr Leu Ala Phe Lys Gln Asn Asn Lys Ile Ala Leu Ser Phe Asn Ala 115 120 125	3621
GAA CTT AAT TTT ACG GGG TGG CGA GGT ATT GCT GTT CCT TTT CGT GAT Glu Leu Asn Phe Thr Gly Trp Arg Gly Ile Ala Val Pro Phe Arg Asp 130 135 140	3669
ATG CAA GGC TCT GCG ACA GGT CAA CTT GAT CAA TTA GTG ATC ACC GCT Met Gln Gly Ser Ala Thr Gly Gln Leu Asp Gln Leu Val Ile Thr Ala 145 150 155 160	3717
CCA AAC CAA GCC GGA ACA CTC TTT TTT GAT CAA ATC ATC ATG AGT GTA Pro Asn Gln Ala Gly Thr Leu Phe Phe Asp Gln Ile Ile Met Ser Val 165 170 175	3765
CCG TTA GAC AAT CGT TGG GCA GTA CCT GAC TAT CAA ACA CCT TAC GTA Pro Leu Asp Asn Arg Trp Ala Val Pro Asp Tyr Gln Thr Pro Tyr Val 180 185 190	3813
AAT AAC GCA GTA AAC ACG ATG GTT AGT AAA AAC TGG AGT GCA TTA TTG Asn Asn Ala Val Asn Thr Met Val Ser Lys Asn Trp Ser Ala Leu Leu 195 200 205	3861

SUBSTITUTE SHEET (RULE 26)

- 131 -

ATG TAC GAT CAG ATG TTT CAA GCC CAT TAC CCT ACT TTA AAC TTC GAT Met Tyr Asp Gln Met Phe Gln Ala His Tyr Pro Thr Leu Asn Phe Asp 210 215 220	3909
ACT GAA TTT CGC GAT GAC CAA ACA GAA ATG GCT TCG ATT TAT CAG CGC Thr Glu Phe Arg Asp Asp Gln Thr Glu Met Ala Ser Ile Tyr Gln Arg 225 230 235 240	3957
TTT GAA TAT TAT CAA GGA ATT CGT AGT GAT AAA AAA ATT ACT CCA GAT Phe Glu Tyr Tyr Gln Gly Ile Arg Ser Asp Lys Lys Ile Thr Pro Asp 245 250 255	4005
ATG CTA GAT AAA CAT TTA GCA TTA TGG GAA AAA TTG GTG TTA ACA CAA Met Leu Asp Lys His Leu Ala Leu Trp Glu Lys Leu Val Leu Thr Gln 260 265 270	4053
CAC GCT GAT GGC TCA ATC ACA GGA AAA GCC CTT GAT CAC CCT AAC CGG His Ala Asp Gly Ser Ile Thr Gly Lys Ala Leu Asp His Pro Asn Arg 275 280 285	4101
CAA CAT TTT ATG AAA GTC GAA GGT GTA TTT AGT GAG GGG ACT CAA AAA Gln His Phe Met Lys Val Glu Gly Val Phe Ser Glu Gly Thr Gln Lys 290 295 300	4149
GCA TTA CTT GAT GCC AAT ATG CTA AGA GAT GTG GGC AAA ACG CTT CTT Ala Leu Leu Asp Ala Asn Met Leu Arg Asp Val Gly Lys Thr Leu Leu 305 310 315 320	4197
CAA ACT GCT ATT TAC TTG CGT AGC GAT TCA TTA TCA GCA ACT GAT AGA Gln Thr Ala Ile Tyr Leu Arg Ser Asp Ser Leu Ser Ala Thr Asp Arg 325 330 335	4245
AAA AAA TTA GAA GAG CGC TAT TTA TTA GGT ACT CGT TAT GTC CTT GAA Lys Lys Leu Glu Glu Arg Tyr Leu Leu Gly Thr Arg Tyr Val Leu Glu 340 345 350	4293
CAA GGT TTT ACA CGA GGA AGT GGT TAT CAA ATT ATT ACT CAT GTT GGT Gln Gly Phe Thr Arg Gly Ser Gly Tyr Gln Ile Ile Thr His Val Gly 355 360 365	4341
TAC CAA ACC AGA GAA CTT TTT GAT GCA TGG TTT ATT GGC CGT CAT GTT Tyr Gln Thr Arg Glu Leu Phe Asp Ala Trp Phe Ile Gly Arg His Val 370 375 380	4389
CTT GCA AAA AAT AAC CTT TTA GCC CCC ACT CAA CAA GCT ATG ATG TGG Leu Ala Lys Asn Asn Leu Leu Ala Pro Thr Gln Gln Ala Met Met Trp 385 390 395 400	4437
TAC AAC GCC ACA GGA CGT ATT TTT GAA AAA AAT AAT GAA ATT GTT GAT Tyr Asn Ala Thr Gly Arg Ile Phe Glu Lys Asn Asn Glu Ile Val Asp 405 410 415	4485
GCA AAT GTC GAT ATT CTC AAT ACT CAA TTG CAA TGG ATG ATA AAA AGC Ala Asn Val Asp Ile Leu Asn Thr Gln Leu Gln Trp Met Ile Lys Ser 420 425 430	4533
TTA TTG ATG CTA CCG GAT TAT CAA CAA CGT CAA CAA GCC TTA GCG CAA Leu Leu Met Leu Pro Asp Tyr Gln Gln Arg Gln Gln Ala Leu Ala Gln 435 440 445	4581
CTG CAA AGT TGG CTA AAT AAA ACC ATT CTA AGC TCA AAA GGT GTT GCT Leu Gln Ser Trp Leu Asn Lys Thr Ile Leu Ser Ser Lys Gly Val Ala	4629

SUBSTITUTE SHEET (RULE 26)

- 132 -

450	455	460	
GGC GGT TTC AAA TCT GAT GGT TCT ATT TTT CAC CAT TCA CAA CAT TAC Gly Gly Phe Lys Ser Asp Gly Ser Ile Phe His His Ser Gln His Tyr 465 470 475 480			4677
CCC GCT TAT GCT AAA GAT GCA TTT GGT GGT TTA GCA CCC AGT GTT TAT Pro Ala Tyr Ala Lys Asp Ala Phe Gly Gly Leu Ala Pro Ser Val Tyr 485 490 495			4725
GCA TTA AGT GAT TCA CCT TTT CGC TTA TCT ACT TCA GCA CAT GAG CGT Ala Leu Ser Asp Ser Pro Phe Arg Leu Ser Thr Ser Ala His Glu Arg 500 505 510			4773
TTA AAA GAT GTT TTG TTA AAA ATG CGG ATC TAC ACC AAA GAG ACA CAA Leu Lys Asp Val Leu Leu Lys Met Arg Ile Tyr Thr Lys Glu Thr Gln 515 520 525			4821
ATT CCT GTG GTA TTA AGT GGT CGT CAT CCA ACT GGG TTG CAT AAA ATA Ile Pro Val Val Leu Ser Gly Arg His Pro Thr Gly Leu His Lys Ile 530 535 540			4869
GGG ATC GCG CCA TTT AAA TGG ATG GCA TTA GCA GGA ACC CCA GAT GGC Gly Ile Ala Pro Phe Lys Trp Met Ala Leu Ala Gly Thr Pro Asp Gly 545 550 555 560			4917
AAA CAA AAG TTA GAT ACC ACA TTA TCC GCC GCT TAT GCA AAA TTA GAC Lys Gln Lys Leu Asp Thr Thr Leu Ser Ala Ala Tyr Ala Lys Leu Asp 565 570 575			4965
AAC AAA ACG CAT TTT GAA GGC ATT AAC GCT GAA AGT GAG CCA GTC GGC Asn Lys Thr His Phe Glu Gly Ile Asn Ala Glu Ser Glu Pro Val Gly 580 585 590			5013
GCA TGG GCA ATG AAT TAT GCA TCA ATG GCA ATA CAA CGA AGA GCA TCG Ala Trp Ala Met Asn Tyr Ala Ser Met Ala Ile Gln Arg Arg Ala Ser 595 600 605			5061
ACC CAA TCA CCA CAA CAA AGC TGG CTC GCC ATA GCG CGC GGT TTT AGC Thr Gln Ser Pro Gln Gln Ser Trp Leu Ala Ile Ala Arg Gly Phe Ser 610 615 620			5109
CGT TAT CTT GTT GGT AAT GAA AGC TAT GAA AAT AAC AAC CGT TAT GGT Arg Tyr Leu Val Gly Asn Glu Ser Tyr Glu Asn Asn Asn Arg Tyr Gly 625 630 635 640			5157
CGT TAT TTA CAA TAT GGA CAA TTG GAA ATT ATT CCA GCT GAT TTA ACT Arg Tyr Leu Gln Tyr Gly Gln Leu Glu Ile Ile Pro Ala Asp Leu Thr 645 650 655			5205
CAA TCA GGG TTT AGC CAT GCT GGA TGG GAT TGG AAT AGA TAT CCA GGT Gln Ser Gly Phe Ser His Ala Gly Trp Asp Trp Asn Arg Tyr Pro Gly 660 665 670			5253
ACA ACA ACT ATT CAT CTT CCC TAT AAC GAA CTT GAA GCA AAA CTT AAT Thr Thr Thr Ile His Leu Pro Tyr Asn Glu Leu Glu Ala Lys Leu Asn 675 680 685			5301
CAA TTA CCT GCT GCA GGT ATT GAA GAA ATG TTG CTT TCA ACA GAA AGT Gln Leu Pro Ala Ala Gly Ile Glu Glu Met Leu Ser Thr Glu Ser 690 695 700			5349

SUBSTITUTE SHEET (RULE 26)

- 133 -

TAC	TCT	GGT	GCA	AAT	ACC	CTT	AAT	AAT	AAC	AGT	ATG	TTT	GCC	ATG	AAA	5397
Tyr	Ser	Gly	Ala	Asn	Thr	Leu	Asn	Asn	Asn	Ser	Met	Phe	Ala	Met	Lys	
705					710					715					720	
TTA	CAC	GGT	CAC	AGT	AAA	TAT	CAA	CAA	CAA	AGC	TTA	AGG	GCA	AAT	AAA	5445
Leu	His	Gly	His	Ser	Lys	Tyr	Gln	Gln	Gln	Ser	Leu	Arg	Ala	Asn	Lys	
				725					730						735	
TCC	TAT	TTC	TTA	TTT	GAT	AAT	AGA	GTT	ATT	GCT	TTA	GGC	TCA	GGT	ATT	5493
Ser	Tyr	Phe	Leu	Phe	Asp	Asn	Arg	Val	Ile	Ala	Leu	Gly	Ser	Gly	Ile	
			740					745					750			
GAA	AAT	GAT	GAT	AAA	CAA	CAT	ACG	ACC	GAA	ACA	ACA	CTA	TTC	CAG	TTT	5541
Glu	Asn	Asp	Asp	Lys	Gln	His	Thr	Thr	Glu	Thr	Thr	Leu	Phe	Gln	Phe	
		755					760					765				
GCC	GTC	CCT	AAA	TTA	CAG	TCA	GTG	ATC	ATT	AAT	GGC	AAA	AAG	GTA	AAT	5589
Ala	Val	Pro	Lys	Leu	Gln	Ser	Val	Ile	Ile	Asn	Gly	Lys	Lys	Val	Asn	
	770					775					780					
CAA	TTA	GAT	ACT	CAA	TTA	ACT	TTA	AAT	AAT	GCA	GAT	ACA	TTA	ATT	GAT	5637
Gln	Leu	Asp	Thr	Gln	Leu	Thr	Leu	Asn	Asn	Ala	Asp	Thr	Leu	Ile	Asp	
785					790					795					800	
CCT	GCC	GGC	AAT	TTA	TAT	AAG	CTC	ACT	AAA	GGA	CAA	ACT	GTA	AAA	TTT	5685
Pro	Ala	Gly	Asn	Leu	Tyr	Lys	Leu	Thr	Lys	Gly	Gln	Thr	Val	Lys	Phe	
			805						810					815		
AGT	TAT	CAA	AAA	CAA	CAT	TCA	CTT	GAT	GAT	AGA	AAT	TCA	AAA	CCA	ACA	5733
Ser	Tyr	Gln	Lys	Gln	His	Ser	Leu	Asp	Asp	Arg	Asn	Ser	Lys	Pro	Thr	
			820					825					830			
GAA	CAA	TTA	TTT	GCA	ACA	GCT	GTT	ATT	TCT	CAT	GGT	AAG	GCA	CCG	AGT	5781
Glu	Gln	Leu	Phe	Ala	Thr	Ala	Val	Ile	Ser	His	Gly	Lys	Ala	Pro	Ser	
		835					840					845				
AAT	GAA	AAT	TAT	GAA	TAT	GCA	ATA	GCT	ATC	GAA	GCA	CAA	AAT	AAT	AAA	5829
Asn	Glu	Asn	Tyr	Glu	Tyr	Ala	Ile	Ala	Ile	Glu	Ala	Gln	Asn	Asn	Lys	
	850					855					860					
GCT	CCC	GAA	TAC	ACA	GTA	TTA	CAA	CAT	AAT	GAT	CAG	CTC	CAT	GCG	GTA	5877
Ala	Pro	Glu	Tyr	Thr	Val	Leu	Gln	His	Asn	Asp	Gln	Leu	His	Ala	Val	
865					870					875					880	
AAA	GAT	AAA	ATA	ACC	CAA	GAA	GAG	GGA	TAT	GCT	TTT	TTT	GAA	GCC	ACT	5925
Lys	Asp	Lys	Ile	Thr	Gln	Glu	Glu	Gly	Tyr	Ala	Phe	Phe	Glu	Ala	Thr	
				885					890					895		
AAG	TTA	AAA	TCA	GCG	GAT	GCA	ACA	TTA	TTA	TCC	AGT	GAT	GCG	CCG	GTT	5973
Lys	Leu	Lys	Ser	Ala	Asp	Ala	Thr	Leu	Leu	Ser	Ser	Asp	Ala	Pro	Val	
			900					905					910			
ATG	GTC	ATG	GCT	AAA	ATA	CAA	AAT	CAG	CAA	TTA	ACA	TTA	AGT	ATT	GTT	6021
Met	Val	Met	Ala	Lys	Ile	Gln	Asn	Gln	Gln	Leu	Thr	Leu	Ser	Ile	Val	
		915						920					925			
AAT	CCT	GAT	TTA	AAT	TTA	TAT	CAA	GGT	AGA	GAA	AAA	GAT	CAA	TTT	GAT	6069
Asn	Pro	Asp	Leu	Asn	Leu	Tyr	Gln	Gly	Arg	Glu	Lys	Asp	Gln	Phe	Asp	
	930					935					940					
GAT	AAA	GGT	AAT	CAA	ATC	GAA	GTT	AGT	GTT	TAT	TCT	CGT	CAT	TGG	CTT	6117
Asp	Lys	Gly	Asn	Gln	Ile	Glu	Val	Ser	Val	Tyr	Ser	Arg	His	Trp	Leu	

SUBSTITUTE SHEET (RULE 26)

- 134 -

945	950	955	960	
ACA GCA GAA TCG CAA TCA ACA AAT AGT ACT ATT ACC GTA AAA GGA ATA				6165
Thr Ala Glu Ser Gln Ser Thr Asn Ser Thr Ile Thr Val Lys Gly Ile	965	970	975	
TGG AAA TTA ACG ACA CCT CAA CCC GGT GTT ATT ATT AAG CAC CAC AAT				6213
Trp Lys Leu Thr Thr Pro Gln Pro Gly Val Ile Ile Lys His His Asn	980	985	990	
AAC AAC ACT CTT ATT ACG ACA ACA ACC ATA CAG GCA ACA CCT ACT GTT				6261
Asn Asn Thr Leu Ile Thr Thr Thr Thr Ile Gln Ala Thr Pro Thr Val	995	1000	1005	
ATT AAT TTA GTT AAG TAAATTTTCGT AACTTTTAAA CTAAAGAGTC TCGACATAAA				6316
Ile Asn Leu Val Lys	1010			
AATATCGAGA CTCTTTTTTAT TAAAAAATTA AAAACAAGTT AACGAATGAA TTAATTATTT				6376
GAAAAATAAA AAATAAATCG ATAGCTTTTAT TATTGATAAT AAATGTGTTG TGCTCAATGG				6436
TTATTTTGTGTT ATTCTCTGCG CGGATGCTTG GATCAATCTG GTTCAAGCAT ATCGCAAGCA				6496
CCAGAACGAA AAAAGCCCCG GGT				6519

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1013 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Met	Leu	Ile	Lys	Asn	Pro	Leu	Ala	His	Ala	Val	Thr	Leu	Ser	Leu	Cys
1				5						10				15	
Leu	Ser	Leu	Pro	Ala	Gln	Ala	Leu	Pro	Thr	Leu	Ser	His	Glu	Ala	Phe
			20					25					30		
Gly	Asp	Ile	Tyr	Leu	Phe	Glu	Gly	Glu	Leu	Pro	Asn	Thr	Leu	Thr	Thr
		35					40					45			
Ser	Asn	Asn	Asn	Gln	Leu	Ser	Leu	Ser	Lys	Gln	His	Ala	Lys	Asp	Gly
	50					55					60				
Glu	Gln	Ser	Leu	Lys	Trp	Gln	Tyr	Gln	Pro	Gln	Ala	Thr	Leu	Thr	Leu
65					70				75						80
Asn	Asn	Ile	Val	Asn	Tyr	Gln	Asp	Asp	Lys	Asn	Thr	Ala	Thr	Pro	Leu
				85					90					95	
Thr	Phe	Met	Met	Trp	Ile	Tyr	Asn	Glu	Lys	Pro	Gln	Ser	Ser	Pro	Leu
			100					105					110		
Thr	Leu	Ala	Phe	Lys	Gln	Asn	Asn	Lys	Ile	Ala	Leu	Ser	Phe	Asn	Ala
		115					120					125			

SUBSTITUTE SHEET (RULE 26)

- 135 -

Glu Leu Asn Phe Thr Gly Trp Arg Gly Ile Ala Val Pro Phe Arg Asp
 130 135 140
 Met Gln Gly Ser Ala Thr Gly Gln Leu Asp Gln Leu Val Ile Thr Ala
 145 150 155 160
 Pro Asn Gln Ala Gly Thr Leu Phe Phe Asp Gln Ile Ile Met Ser Val
 165 170 175
 Pro Leu Asp Asn Arg Trp Ala Val Pro Asp Tyr Gln Thr Pro Tyr Val
 180 185 190
 Asn Asn Ala Val Asn Thr Met Val Ser Lys Asn Trp Ser Ala Leu Leu
 195 200 205
 Met Tyr Asp Gln Met Phe Gln Ala His Tyr Pro Thr Leu Asn Phe Asp
 210 215 220
 Thr Glu Phe Arg Asp Asp Gln Thr Glu Met Ala Ser Ile Tyr Gln Arg
 225 230 235 240
 Phe Glu Tyr Tyr Gln Gly Ile Arg Ser Asp Lys Lys Ile Thr Pro Asp
 245 250 255
 Met Leu Asp Lys His Leu Ala Leu Trp Glu Lys Leu Val Leu Thr Gln
 260 265 270
 His Ala Asp Gly Ser Ile Thr Gly Lys Ala Leu Asp His Pro Asn Arg
 275 280 285
 Gln His Phe Met Lys Val Glu Gly Val Phe Ser Glu Gly Thr Gln Lys
 290 295 300
 Ala Leu Leu Asp Ala Asn Met Leu Arg Asp Val Gly Lys Thr Leu Leu
 305 310 315 320
 Gln Thr Ala Ile Tyr Leu Arg Ser Asp Ser Leu Ser Ala Thr Asp Arg
 325 330 335
 Lys Lys Leu Glu Glu Arg Tyr Leu Leu Gly Thr Arg Tyr Val Leu Glu
 340 345 350
 Gln Gly Phe Thr Arg Gly Ser Gly Tyr Gln Ile Ile Thr His Val Gly
 355 360 365
 Tyr Gln Thr Arg Glu Leu Phe Asp Ala Trp Phe Ile Gly Arg His Val
 370 375 380
 Leu Ala Lys Asn Asn Leu Leu Ala Pro Thr Gln Gln Ala Met Met Trp
 385 390 395 400
 Tyr Asn Ala Thr Gly Arg Ile Phe Glu Lys Asn Asn Glu Ile Val Asp
 405 410 415
 Ala Asn Val Asp Ile Leu Asn Thr Gln Leu Gln Trp Met Ile Lys Ser
 420 425 430
 Leu Leu Met Leu Pro Asp Tyr Gln Gln Arg Gln Gln Ala Leu Ala Gln
 435 440 445
 Leu Gln Ser Trp Leu Asn Lys Thr Ile Leu Ser Ser Lys Gly Val Ala
 450 455 460

SUBSTITUTE SHEET (RULE 26)

- 136 -

Gly	Gly	Phe	Lys	Ser	Asp	Gly	Ser	Ile	Phe	His	His	Ser	Gln	His	Tyr	465	470	475	480
Pro	Ala	Tyr	Ala	Lys	Asp	Ala	Phe	Gly	Gly	Leu	Ala	Pro	Ser	Val	Tyr	485	490	495	
Ala	Leu	Ser	Asp	Ser	Pro	Phe	Arg	Leu	Ser	Thr	Ser	Ala	His	Glu	Arg	500	505	510	
Leu	Lys	Asp	Val	Leu	Leu	Lys	Met	Arg	Ile	Tyr	Thr	Lys	Glu	Thr	Gln	515	520	525	
Ile	Pro	Val	Val	Leu	Ser	Gly	Arg	His	Pro	Thr	Gly	Leu	His	Lys	Ile	530	535	540	
Gly	Ile	Ala	Pro	Phe	Lys	Trp	Met	Ala	Leu	Ala	Gly	Thr	Pro	Asp	Gly	545	550	555	560
Lys	Gln	Lys	Leu	Asp	Thr	Thr	Leu	Ser	Ala	Ala	Tyr	Ala	Lys	Leu	Asp	565	570	575	
Asn	Lys	Thr	His	Phe	Glu	Gly	Ile	Asn	Ala	Glu	Ser	Glu	Pro	Val	Gly	580	585	590	
Ala	Trp	Ala	Met	Asn	Tyr	Ala	Ser	Met	Ala	Ile	Gln	Arg	Arg	Ala	Ser	595	600	605	
Thr	Gln	Ser	Pro	Gln	Gln	Ser	Trp	Leu	Ala	Ile	Ala	Arg	Gly	Phe	Ser	610	615	620	
Arg	Tyr	Leu	Val	Gly	Asn	Glu	Ser	Tyr	Glu	Asn	Asn	Asn	Arg	Tyr	Gly	625	630	635	640
Arg	Tyr	Leu	Gln	Tyr	Gly	Gln	Leu	Glu	Ile	Ile	Pro	Ala	Asp	Leu	Thr	645	650	655	
Gln	Ser	Gly	Phe	Ser	His	Ala	Gly	Trp	Asp	Trp	Asn	Arg	Tyr	Pro	Gly	660	665	670	
Thr	Thr	Thr	Ile	His	Leu	Pro	Tyr	Asn	Glu	Leu	Glu	Ala	Lys	Leu	Asn	675	680	685	
Gln	Leu	Pro	Ala	Ala	Gly	Ile	Glu	Glu	Met	Leu	Leu	Ser	Thr	Glu	Ser	690	695	700	
Tyr	Ser	Gly	Ala	Asn	Thr	Leu	Asn	Asn	Asn	Ser	Met	Phe	Ala	Met	Lys	705	710	715	720
Leu	His	Gly	His	Ser	Lys	Tyr	Gln	Gln	Gln	Ser	Leu	Arg	Ala	Asn	Lys	725	730	735	
Ser	Tyr	Phe	Leu	Phe	Asp	Asn	Arg	Val	Ile	Ala	Leu	Gly	Ser	Gly	Ile	740	745	750	
Glu	Asn	Asp	Asp	Lys	Gln	His	Thr	Glu	Thr	Thr	Leu	Phe	Gln	Phe		755	760	765	
Ala	Val	Pro	Lys	Leu	Gln	Ser	Val	Ile	Ile	Asn	Gly	Lys	Lys	Val	Asn	770	775	780	
Gln	Leu	Asp	Thr	Gln	Leu	Thr	Leu	Asn	Asn	Ala	Asp	Thr	Leu	Ile	Asp	785	790	795	800

SUBSTITUTE SHEET (RULE 26)

- 137 -

[illegible]

(2) INFORMATION FOR SEO ID NO:41:

- ```
(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO
```

(xi) SEQUENCE DESCRIPTION: SEO ID NO:41:

ATTTCAGGA AATCTGCATA TGCTAATAAA AAACCC

- 138 -

What is claimed is:

1. A purified isolated DNA fragment of Proteus vulgaris (P. vulgaris) comprising a sequence encoding for the chondroitinase I enzyme.
2. A purified isolated DNA fragment of P. vulgaris, wherein the fragment comprises a sequence which hybridizes with a nucleic acid sequence encoding for the amino acids numbered 1-1021 of SEQ ID NO:2 or a biological equivalent thereof.
3. The purified isolated DNA fragment of Claim 2, wherein the fragment has the sequence of (a) the nucleotides numbered 119-3181 of SEQ ID NO:1, or (b) the nucleotides numbered 119-3181 of SEQ ID NO:3.
4. A purified isolated DNA fragment of P. vulgaris, wherein the fragment comprises a sequence which hybridizes with a nucleic acid sequence encoding for the amino acids numbered 25-1021 of SEQ ID NO:2 or a biological equivalent thereof.
5. The purified isolated DNA fragment of Claim 4, wherein the fragment has the sequence of (a) the nucleotides numbered 191-3181 of SEQ ID NO:1, or (b) the nucleotides numbered 191-3181 of SEQ ID NO:3.
6. A purified isolated DNA fragment of P. vulgaris, wherein the fragment comprises a sequence which hybridizes with a nucleic acid sequence encoding for the amino acids numbered 24-1021 of SEQ ID NO:5 or a biological equivalent thereof.
7. The purified isolated DNA fragment of Claim 6, wherein the fragment has the sequence of nucleotides numbered 188-3181 of SEQ ID NO:4.
8. A purified isolated DNA fragment of Proteus vulgaris (P. vulgaris) comprising a sequence encoding for chondroitinase II enzyme.
9. A purified isolated DNA fragment of P.



- 139 -

vulgaris, wherein the fragment comprises a sequence which hybridizes with a nucleic acid sequence encoding for (a) the amino acids numbered 1-1013 of SEQ ID NO:40 or a biological equivalent thereof, or (b) the amino acids numbered 24-1013 of SEQ ID NO:40 or a biological equivalent thereof.

10. The purified isolated DNA fragment of Claim 9, wherein the fragment has the sequence of nucleotides (a) numbered 3238-6276 of SEQ ID NO:39, or (b) numbered 3307-6276 of SEQ ID NO:39.

11. A plasmid containing a purified isolated DNA fragment of P. vulgaris comprising the sequence of (a) Claim 1 or (b) Claim 8.

12. The plasmid of Claim 11 wherein the plasmid is that designated pTM49-6 or that designated LP<sup>2</sup>1359.

13. A host cell transformed with the plasmid of Claim 11.

14. The host cell of Claim 13 wherein the plasmid is that designated pTM49-6 (ATCC 69234) or that designated LP<sup>2</sup>1359 (ATCC 69598).

15. A purified isolated recombinant chondroitinase I enzyme.

16. The chondroitinase I enzyme of Claim 15, whose amino acid sequence is depicted for (a) the amino acids numbered 1-1021 of SEQ ID NO:2 or a biological equivalent thereof, (b) the amino acids numbered 25-1021 of SEQ ID NO:2 or a biological equivalent thereof, or (c) the amino acids numbered 24-1021 of SEQ ID NO:5 or a biological equivalent thereof.

17. A purified isolated recombinant chondroitinase II enzyme.

18. The chondroitinase II enzyme of Claim 17, whose amino acid sequence is depicted for (a) the

- 140 -

amino acids numbered 1-1013 of SEQ ID NO:40 or a biological equivalent thereof, or (b) the amino acids numbered 24-1013 of SEQ ID NO:40 or a biological equivalent thereof.

19. A method of producing chondroitinase I enzyme which comprises transforming a host cell with the plasmid of Claim 11(a) and culturing the host cell under conditions which permit expression of said enzyme by the host cell.

20. A method of producing the chondroitinase II enzyme which comprises transforming a host cell with the plasmid of Claim 11(b) and culturing the host cell under conditions which permit expression of said enzyme by the host cell.

21. A method for the isolation and purification of the recombinant chondroitinase I enzyme of Proteus vulgaris from host cells, said method comprising the steps of:

- (a) lysing by homogenization the host cells to release the enzyme into the supernatant;
- (b) subjecting the supernatant to diafiltration to remove salts and other small molecules;
- (c) passing the supernatant through an anion exchange resin-containing column;
- (d) loading the eluate from step (c) to a cation exchange resin-containing column so that the enzyme in the eluate binds to the cation exchange column; and
- (e) eluting the enzyme bound to the cation exchange column with a solvent capable of releasing the enzyme from the column.

22. The method of Claim 21, wherein prior

- 141 -

to step (b), the following two steps are performed:

- (1) treating the supernatant with an acidic solution to precipitate out the enzyme; and
- (2) recovering the pellet and then dissolving it in an alkali solution to again place the enzyme in a basic environment.

23. A recombinant chondroitinase I enzyme isolated and purified by the method of Claim 21 or by the method of Claim 22.

24. A method for the isolation and purification of the recombinant chondroitinase II enzyme of Proteus vulgaris from host cells, said method comprising the steps of:

- (a) lysing by homogenization the host cells to release the enzyme into the supernatant;
- (b) subjecting the supernatant to diafiltration to remove salts and other small molecules;
- (c) passing the supernatant through an anion exchange resin-containing column;
- (d) loading the eluate from step (c) to a cation exchange resin-containing column so that the enzyme in the eluate binds to the cation exchange column; and
- (e) obtaining by affinity elution the enzyme bound to the cation exchange column with a solution of chondroitin sulfate, such that the enzyme is co-eluted with the chondroitin sulfate;
- (f) loading the eluate from step (e) to an anion exchange resin-containing column and eluting the enzyme with a solvent

- 142 -

such that the chondroitin sulfate binds to the column; and

- (g) concentrating the eluate from step (f) and crystallizing out the enzyme from the supernatant which contains an approximately 37 kD contaminant.

25. The method of Claim 24, wherein prior to step (b), the following two steps are performed:

- (1) treating the supernatant with an acidic solution to precipitate out the enzyme; and
- (2) recovering the pellet and then dissolving it in an alkali solution to again place the enzyme in a basic environment.

26. A recombinant chondroitinase II enzyme isolated and purified by the method of Claim 24 or by the method of Claim 25.

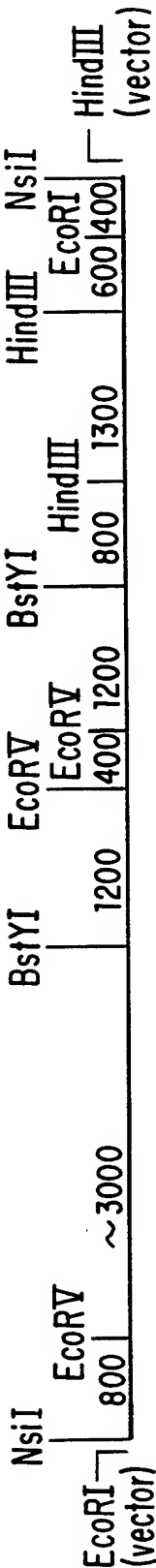


FIG. 1

2/5

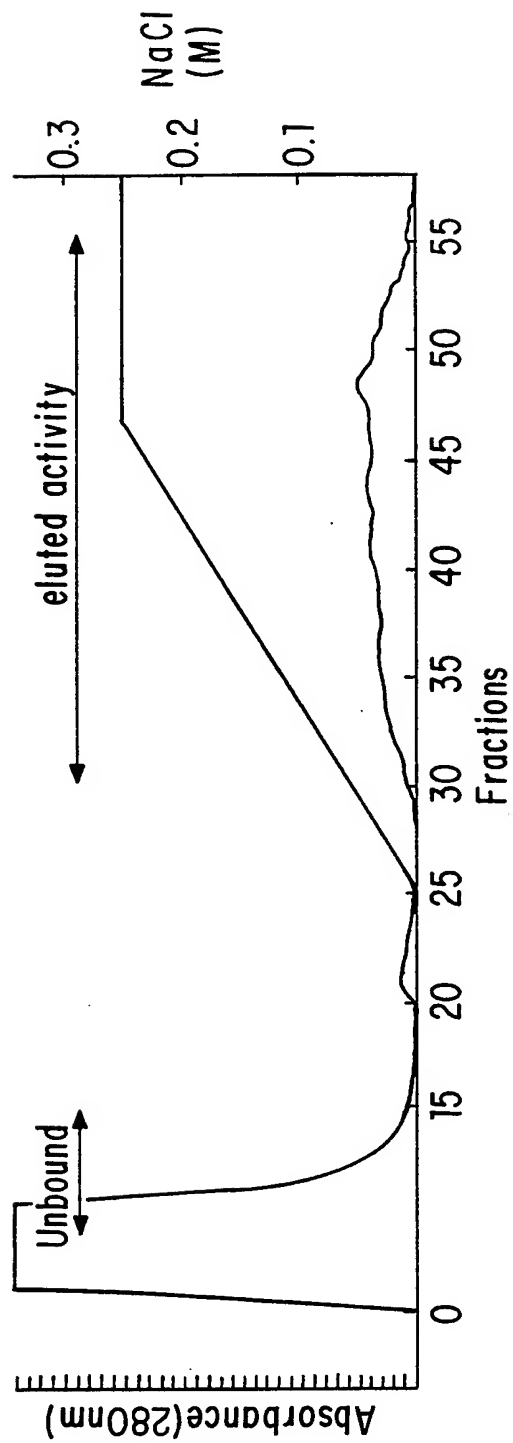


FIG. 2

3/5

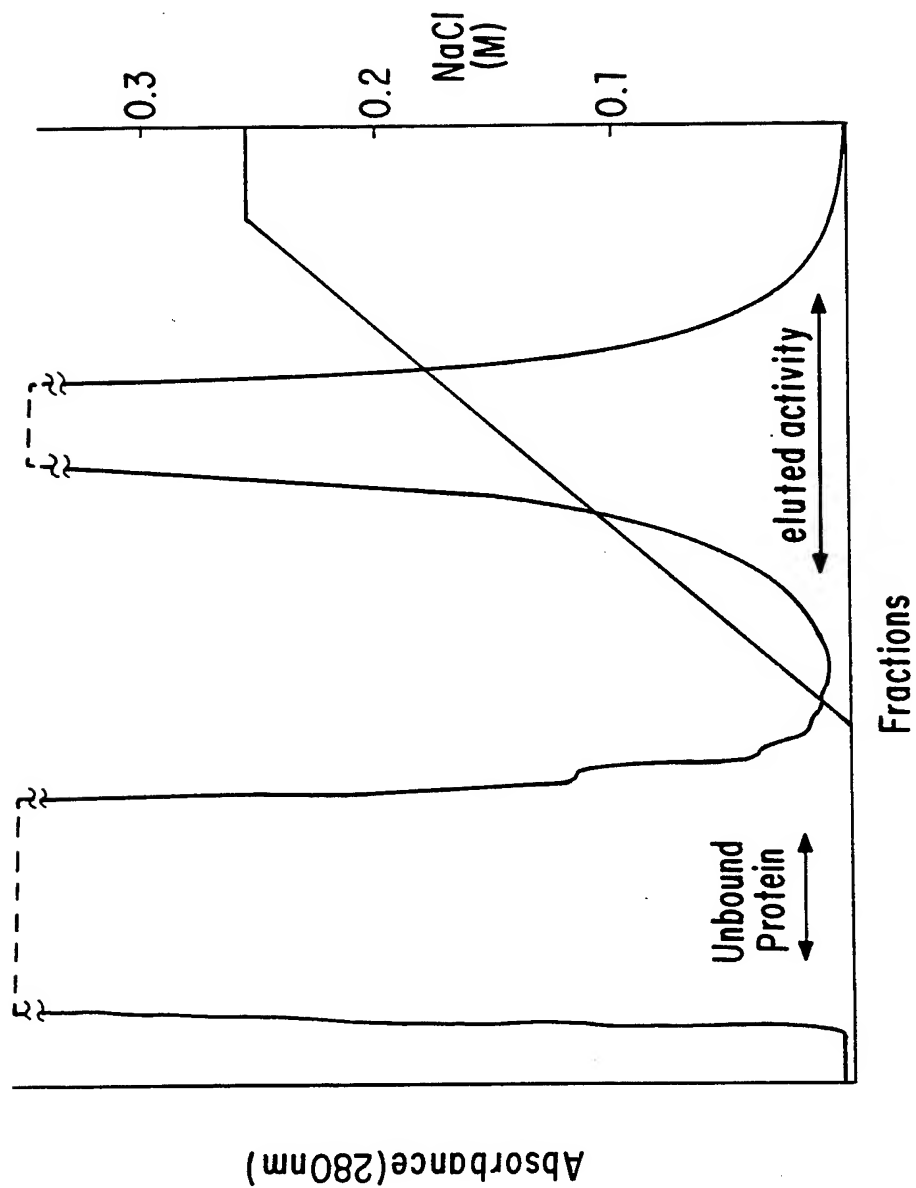


FIG. 3

4/5

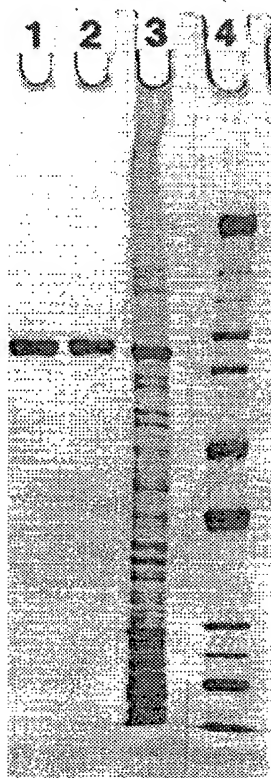


FIG. 4



5/5

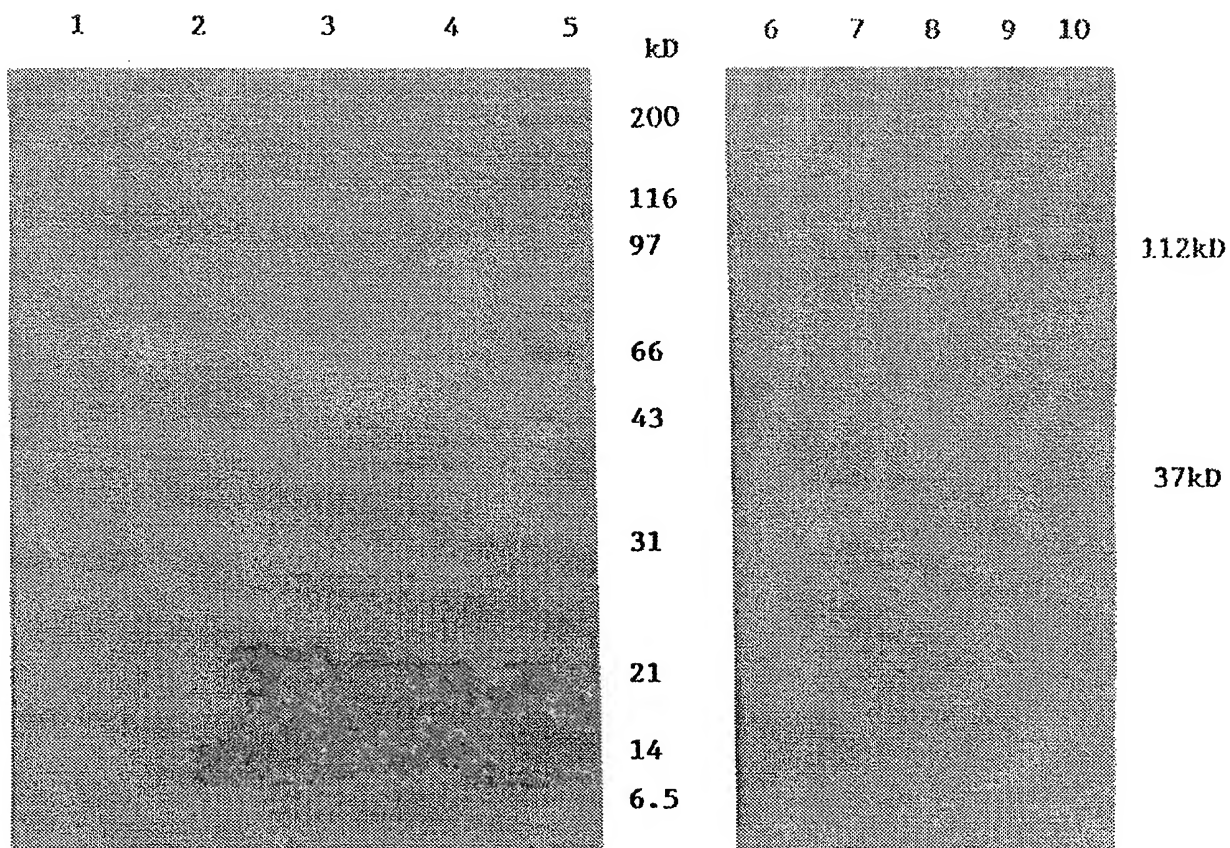


FIG. 5A

FIG. 5B

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/04495

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 1/21, 9/88, 15/31, 15/60, 15/61

US CL : 435/69.1, 232, 252.3, 320.1; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 232, 252.3, 320.1; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CA, BIOSIS, INPADOC, JICST-E search terms: chondroitinase, Proteus vulgaris

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category*      | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                    | Relevant to claim No.                                                |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| X<br>----<br>Y | US, A, 5,049,501 (KATSURAGI et al) 17 September 1991, column 4, line 1 to column 5, line 68.                                                                                                                          | 1,2,4,6,8,<br>9,11,13<br>-----<br>3,5,7,10,<br>12,14,19,<br>20,21-26 |
| X<br>----<br>Y | AGRICULTURAL AND BIOLOGICAL CHEMISTRY, Volume 50, Number 4, issued 1986, Sato et al, "Subunit Structure of Chondroitinase ABC from Proteus vulgaris", pages 1057-1059, see page 1057, paragraph bridging columns 1-2. | 15-18<br>-----<br>21-26                                              |

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\*

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\*

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*&\*

document member of the same patent family

Date of the actual completion of the international search

01 JULY 1994

Date of mailing of the international search report

JUL 15 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

STEPHEN WALSH

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/04495

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category*       | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                             | Relevant to claim No.   |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
| X<br>-----<br>Y | THE JOURNAL OF BIOLOGICAL CHEMISTRY, Volume 243, Number 7, issued 10 April 1968, Yamagata et al, "Purification and Properties of Bacterial Chondroitinases and Chondrosulfatases", pages 1523-1535, see Purification section, pages 1526-1527. | 15-18<br>-----<br>21-26 |
| Y               | US, A, 5,198,355 (KIKUCHI et al) 30 March 1993, see column 3, lines 3-25.                                                                                                                                                                      | 21-26                   |
| Y               | D. M. GLOVER, "DNA CLONING, VOLUME 1, A PRACTICAL APPROACH", published 1985 by IRL Press (Oxford), pages 49-77, see page 49, paragraph 1.                                                                                                      | 11-14,19,<br>20         |

# INTERNATIONAL SEARCH REPORT

In ternational application No.  
PCT/US94/04495

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/04495

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-7, 11-16, 19 and 21-23, directed to DNA encoding chondroitinase I, plasmids, host cells, chondroitinase I, and a method for purifying chondroitinase I, classified in Class 435, subclass 232.
- II. Claims 8-14, 17, 18, 20 and 24-26, directed to DNA encoding chondroitinase II, plasmids, host cells, chondroitinase II, and a method for purifying chondroitinase II, classified in Class 435, subclass 232.

Claims 11-14 link the groups because they include both species of chondroitinase. In the event the additional fee is not paid, claims 11-14 will be searched to the extent they embrace chondroitinase I.

The claims are directed to two distinct species of enzyme which are not so linked by a special technical feature so as to form a single general inventive concept as required by PCT Rules 13.1-13.3. The special technical features of the two species are disclosed as distinguishing them from each other and from the prior art.